Molecular characterization of HPV infection

van der Weele, P.S.J.

2019

document version
Publisher's PDF, also known as Version of record

Molecular characterization of HPV infection

Evaluation of vaccine effects, viral diversity and variant development

Pascal Stefan Johan van der Weele

Molecular characterization of HPV infection
Evaluation of vaccine effects, viral diversity and variant development

ISBN: 978-94-6339-161-0

The work described in this thesis was conducted at the department virology of the vaccination programme at the National Institute for Public Health and the Environment (RIVM) in collaboration with the department of pathology at Amsterdam UMC (location VUmc).

Printing of this thesis was supported by the National Institute for Public Health and the Environment and Amsterdam UMC.

Cover: Eelco van Rooij
Layout and typesetting: Michal Slawinski, thesisprint.eu
Printed in Poland

VRIJE UNIVERSITEIT

MOLECULAR CHARACTERIZATION OF HPV INFECTION

Evaluation of vaccine effects, viral diversity and variant development

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op donderdag 31 oktober 2019 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

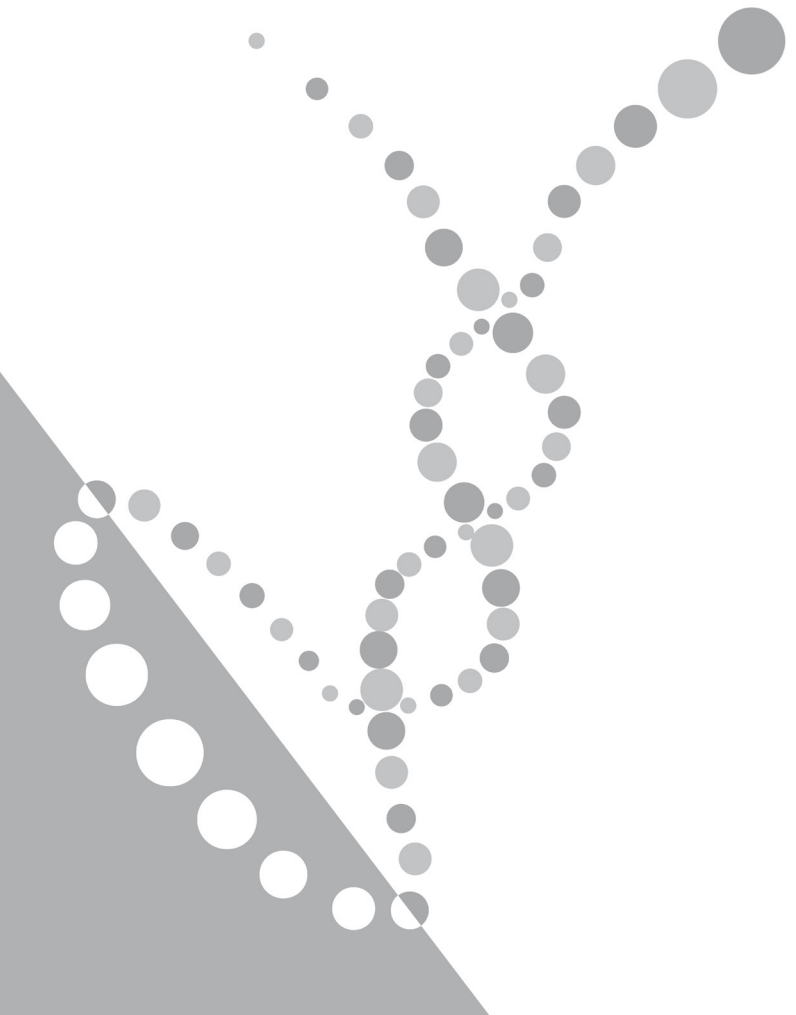
Pascal Stefan Johan van der Weele

geboren te Breda

promotor:	prof.dr. C.J.L.M. Meijer
copromotor:	dr. A.J. King

Table of contents

Chapter 1



General Introduction

Human papillomavirus associated disease and burden

Human papillomavirus (HPV) is the most common sexually transmitted infection worldwide. As a result, approximately 80% of the world population will, at some point in their life, contract a HPV infection [1]. Worldwide, 8.0% of all cancers are cancers of the cervix [2], for which HPV has been recognized as a necessary, but insufficient cause [3, 4]. Besides cervical cancer in women, HPV has also been associated with cancers of the anus, vulva, penis, and various cancers in the head-and-neck region. Table 1 summarizes worldwide HPV related cancer incidence in 2012 [2], supplemented with data from 2018 where available. Worth noting are the differences in prevalence of different HPV types in different areas of the world [5, 6].

Table 1: Worldwide number of cancer cases attributable to HPV and corresponding attributable fraction (AF) in 2012 by cancer site, sex and age. Adapted from de Martel *et al.* [2] and supplemented with numbers from Globocan 2018 where available.

HPV-related cancer site (ICD-10 code)	Number of incident cases[a, b, c]	Number attributable to HPV	AF (%)[d]	Number attributable to HPV by gender		Number attributable to HPV by age group		
				Males	Females	< 50 years	50–69 years	70+ years
Cervix uteri[b] (C53)	570,000	570,000	100	0	570,000	250,000	250,000	71,000
Anus[e] (C21)	40,000	35,000	88	17,000	18,000	6,600	17,000	12,000
Vulva[b] (C51)	44,000	11,000	24.9	0	11,000	2,100	4,000	5,000
Vagina[b] (C52)	18,000	14,000	78	0	14,000	2,800	6,200	4,700
Penis[b] (C60)	34,000	17,000	50	17,000	0	2,000	8,300	5,900
Oropharynx[e] (C01, C09–10)	96,000	29,000	30.8	24,000	5,500	5,400	18,000	6,000
Oral cavity[e] (C02–06)	200,000	4,400	2.2	2,900	1,500	890	2,300	1,200
Larynx[b] (C32)	180,000	4,300	2.4	3,700	540	490	2,500	1,200
Total HPV-related sites	1,200,000	680,000		65,000	620,000	270,000	310,000	110,000

[a]Source: Globocan 2012.
[b]Source: Globocan 2018
[c]Numbers are rounded to two significant digits.
[d]Attributable fractions according to de Martel, 2012 were used for Globocan 2018 data.
[e]These cancer sites were not directly available in GLOBOCAN 2012; therefore, data from the Cancer Incidence in Five Continents (CI5-X) database were used to estimate the corresponding number of cases.

Papillomavirus family properties

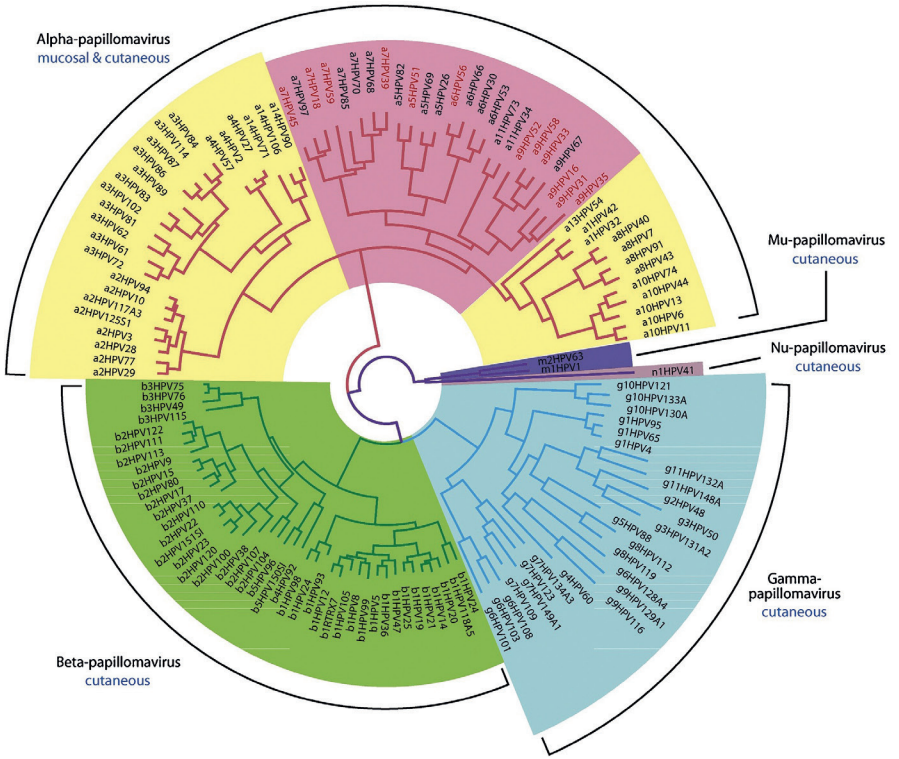
Papillomaviruses are double-stranded, circular DNA viruses of around 7.9kb which exclusively infect cutaneous and mucosal epithelia. To date, over 200 unique HPV types have been characterized, which are structured into five genera (Alpha, Beta, Gamma, Mu and Nu) [7–9]. HPV types are considered to be either low-risk (lrHPV) or high-risk (hrHPV) based on their phylogenetic and epidemiological association with oncogenic potential for cervical cancer [10]. Genera and tropisms

of various HPV types is displayed in Figure 1 [11]. HPV types belonging to the Gamma, Mu and Nu genera are typically lrHPV types, which are most commonly associated with palm and plantar warts. Beta-papillomavirus are frequently observed in cutaneous lesions and are associated with the development of skin cancer. Despite this association, they are also considered lrHPV. The hrHPV types are found among the Alpha-papillomavirus genus, which can infect the anogenital tract. The International Agency for Research on Cancer (IARC) maintains a list of carcinogens and their potential risk [10]. Group 1 carcinogens are definitely carcinogenic to humans, while group 2A and 2B contain probable and possible carcinogens respectively. Currently, thirteen HPV types are considered group 1 carcinogens (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66). A further twelve HPV types are considered group 2A or 2B carcinogens (HPV26, 30, 34, 53, 67, 68, 69, 70, 73, 82, 85 and 97) [6, 10, 12]. Within the group of hrHPV types, HPV16 and HPV18 combined cause roughly 70% of all cervical cancer cases worldwide [6], and as a result, have been the most common targets for research and vaccination strategies.

Human papillomavirus genome

HPV types exhibit differences in open reading frames (ORFs) and functional gene expression. The HPV genome is comprised of eight genes, which are split in an 'early' regulatory group (six genes; *E1*, *E2*, *E4*, *E5*, *E6* and *E7*) and a 'late' differentiation related group (two genes; *L1* and *L2*) based on the promoter activity required for transcription of the genes. The HPV genetic layout is shown in Figure 2. After initial infection, the virus exists solely in an episomal state and replicates along with the host DNA. Upon basal cell differentiation, the viral productive phase is initiated, resulting in activation of the 'early' genes. Among the 'early' genes, *E1* encodes a helicase, which can unwind the viral genome. It also recruits cellular factors to the viral origin of replication (ORI) located on the upstream regulatory region (URR) [13]. The E2 protein facilitates early replication of the virus by loading E1 on the viral ORI. E2 also functions as a transcriptional regulator, by downregulating the 'early' phase promoter and as a result inhibiting early expression of E6 and E7 [14].

E6 and *E7* are considered the viral oncogenes and as a result have been intensively studied. The E6 and E7 proteins aid in viral reproduction by modulating cell cycle control by inactivating cell cycle checkpoint control tumor suppressor proteins p53 and pRb (retinoblastoma). As a consequence of pRb inhibition, E2F and related transcription factors are no longer regulated, causing the cell cycle and DNA repair mechanisms to become deregulated, thereby inducing genetic instability [15–17] (Figure 4). The inhibitory effect of E2 on E6 and E7 during early replication initially leads to relatively low viral copy numbers and is therefore an important part of the host immune evasion strategy of the virus [18]. Finally, the E4 and E5 proteins are suggested to function in facilitating replication during differentiation of host cells [19]. During the early phase of viral reproduction, there is no assembly of virus particles yet.



Genus + Species	Type	
Alpha 1	HPV 32	More frequent in mucosal lesions than cutaneous (low risk)
Alpha 8	HPV 7	Oral epithelial hyperplasia, cervical intraepithelial neoplasia
Alpha 10	HPV 6	Condylomata acuminata (HPV 6 and HPV 11)
	HPV 11	Recurrent Respiratory Papillomatosis (HPV 11 and HPV 6)
Alpha 13	HPV 54	Butcher's wart (cutaneous, HPV 7)
Alpha 9	HPV 16	High-risk mucosal lesions Each type shows different carcinogenicity Bowenoid papulosis, Bowen's desease (cutaneous, HPV 16)
	HPV 31	
Alpha 7	HPV 18	
	HPV 45	
Alpha 5	HPV 51	
Alpha 6	HPV 56	
Alpha 11	HPV 34	
Alpha 3		Low-risk mucosal lesions
Alpha14		
Alpha 2	HPV 3	More frequent in cutaneous lesions than mucosal (low-risk)
	HPV 10	Flat warts (Alpha 2)
Alpha 4	HPV 2	Verruca vulgaris (common warts) (Alpha 4)
	HPV 57	
Beta 1	HPV 5	Most frequently causes cutaneous lesions, DNA is also detected in mucosa A possible role in the development of skin cancer
	HPV 8	
Beta 2		Commonly assocciated with Epidermodysplasia verruciformis (EV, Beta 1 and 2) or immunosuppressed patients (aquired EV)
Beta 3		
Beta 4		In general immunocompetent population, infection is asymptomatic DNA is detected on plucked hairs, suggesting the hair follicle as a site of infection
Beta 5		
Gamma 1	HPV 4	Gamma-PV DNA is detected from skin and mucosal sample of healthy population Infection is asymptomatic in general immunocompetent population Causes lesions in immunosuppressed paitients
	HPV 65	
Gamma 4	HPV 60	
other Gammas		Some types are known to cause cutaneous lesions in general population Common warts, Plantar warts, Pigmented warts (HPV 4 and 65), Epidermoid cysts (HPV 60) Histologically distinct intracytoplasmic inclusion bodies
Mu 1	HPV 1	Cutaneous lesion especially in palm and plantar (Myrmecia, HPV 1) Histologically distinct intracytoplasmic inclusion bodies Eccrine duct is thought to be a site of infection
Mu 2	HPV 63	
Nu 1	HPV 41	Isolated from cutaneous lesions

Figure 1: Phylogeny, tropism and pathogenesis of human papillomavirus types (Egawa *et al.*, 2017) [11].

As basal cell differentiation continues the 'late' promoter is induced. This promoter is not subject to E2 regulation and causes high levels of E1, E2, E4 and E5 protein to be expressed. In addition, this promoter leads to expression of the so-called 'late' genes *L1* and *L2*. The L1 and L2 proteins together form the viral capsid. Expression of these proteins leads to virion maturation and assembly in the final stages of basal cell differentiation [19].

Expression of the different viral proteins is tightly regulated and under normal circumstances will not lead to the development of cancer, due to E2 inhibiting expression of *E6* and *E7*. However, in some cases this self-regulation is lost. Most commonly, this occurs following integration of the HPV DNA into the host genome. Integration events often result in disruption of the *E2* gene, causing a loss of inhibitory control over *E6* and *E7* gene expression. Furthermore, integration has been shown to preferentially occur in transcriptionally active regions of the host genome [20]. However, not all cervical cancers contain integrated HPV DNA, suggesting that there are alternatives for *E2* loss of function. Methylation of the E2 binding sites located in the URR of the HPV genome has been shown to cause a loss of E6 and E7 repression without viral integration [21]. Without E6 and E7 regulation by E2, the infected host cells can transform facilitating eventual carcinogenesis.

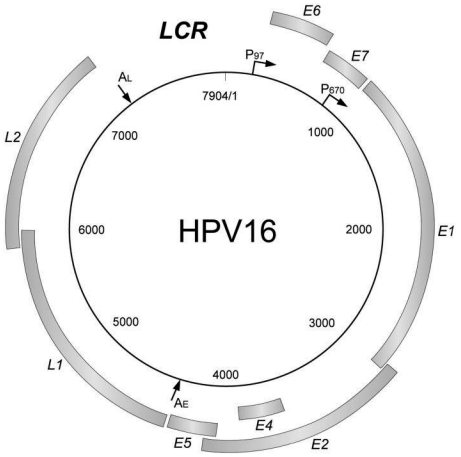


Figure 2: Genome structure of hrHPV with gene locations and functions [22]. The 'early' and 'late' promoters are represented by P97 and P670 respectively. AE and AL denote polyadenylation signals for the 'early' and 'late' promoters respectively.

HPV evolution and nomenclature

Papillomaviruses are a family of host-restricted viruses broadly occurring in the animal kingdom [23]. There are at least 39 genera with unique HPV types, of which only five are known to contain HPV types that can infect humans. With over 200 distinct HPV types described, humans are the best studied host of papillomaviruses [7–9]. Virtually all HPV types are able to establish persistent

infections in their hosts, with a generally benign outcome [24]. PVs are considered to co-evolve with their respective hosts, which has been shown for HPV as well, based on associations between HPV and human genetic diversity [25, 26].

Historically, HPVs are considered different genera if they share <60% similarity on the L1 sequence, as shown in Table2. Within genera, *species* share 60%-70% sequence homology. HPV *sequences* are considered different *types* if they share 70–90% similarity on the L1 genetic sequence [7, 27]. Sequences with >90% homology are considered *intra-typic variants* or *subtypes*. As more whole genome sequence data was being obtained globally, *lineages* and *sublineages* were introduced specifying intra-typic details based on empiric definitions [28–31]. *Lineages* within an HPV *type* differ between 1.0–10% at the whole genome sequence level. If sequences within a *type* differ between 0.5 and 1.0%, they are considered *sublineages*. Sequences differing <0.5% are considered *variants* [28]. Because of the increasing availability of high sensitivity sequencing techniques, new *types* are continuously discovered. A new HPV *type* is acknowledged as such, when it has been cloned and confirmed through re-sequencing by the international HPV reference center (www.hpvcenter.se).

Table 2: Nomenclature of HPV types and subtypes based on [7, 27, 28].

	L1 or full genome	Sequence homology
Genus	L1	<60%
Species	L1	60%-70%
Type	L1	70%-90%
Lineage	Full genome	90.0%-99.0%
Sublineage	Full genome	99.0%-99.5%
Variant	Full genome	>99.5%

Relevance of lineages

For both HPV16 and HPV18, it was found that different viral strains circulated preferentially in specific parts of the world, with geographical distribution shown in Figure 3 [32–34]. In addition, an association was found between geographic heritage of viral strains and ethnic background of the host; HPV16 strains that were specific to a region, were more likely to persist in native human hosts [35]. However, for most other types similar results were not obtained, possibly due to less data being available.

With the initial finding that different HPV16 variants were geographically distributed, it was also found that certain lineages were enriched in cervical cancer specimens [28, 33–39] and specific variants have been associated with predisposition towards persisting infections [40]. In the current era of next-generation sequencing (NGS), it is becoming clearer how HPV16 type-specific variants affect carcinogenicity. Large population level studies allow elucidation of more nuanced effects of sequence differences, while contemporary NGS platforms allow for high throughput

generation of HPV sequence data at relatively low cost [41]. As a result, differential risk for histology-specific outcome measures has been studied for the known HPV16 lineages [34, 42]. Further investigations still are showing that specific conservation of the E7 oncogene is essential for the carcinogenesis of a HPV16 infection [43].

These findings combined illustrate the epidemiological and clinical relevance for investigation of type-specific variants. Most studies investigating lineages do so from a cross-sectional perspective. Longitudinal studies investigating the amount of diversity are far less prevalent and are required to gain an understanding in the development of persistent infections and the amount and origins of diversity of HPV in a study population.

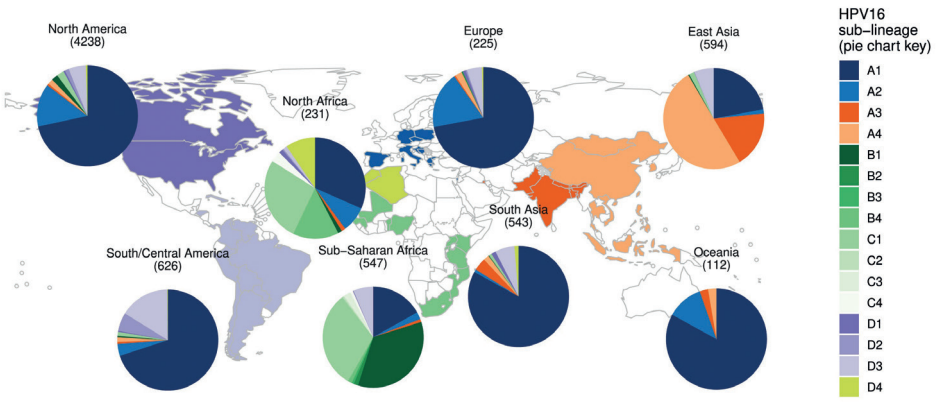


Figure 3: Geographic spread of (sub-)lineages from 7116 HPV16 positive samples which were subjected to whole genome sequencing. Adapted from [34].

Natural history of infection

To gain an understanding of how HPV associated diseases develop, understanding of the biological niche of hrHPVs is required. The cervix connects the vagina and the uterus and is divided in the ectocervix, covering the surface of the vagina, and the endocervix, bordering the endocervical canal of the uterus. The ectocervix is lined with squamous epithelial tissue, while the endocervix is lined with columnar epithelial. Ecto- and endocervix meet at an area called the squamo-columnar junction (SCJ). During puberty, the SCJ shifts from the ectocervix to the endocervix and glandular epithelium is replaced by metaplastic squamous epithelium. The metaplastic epithelial area is called the transformation zone. This metaplastic squamous epithelium is highly susceptible to HPV infection and virtually all HPV -associated cervical lesions originate in this area.

Genital HPV infections commonly develop during sexual intercourse, when micro-abrasions of the cervical epithelium can cause the basal layer to become exposed [44]. Free HPV virions (possibly deposited as a consequence of intercourse) can then reach and enter the basal cells, initiating a HPV infection (Figure 4) [44–46]. Infections can be either transient, meaning clearance will occur within 12–18 months depending on the HPV type, or persistent [47]. It is estimated that around 80% of all HPV infections are transient, while the remaining 20% persist within the host, of which a small subset (1–3.5%) can eventually cause lesions and possibly cervical cancer [1, 48].

Latency of infection

Initially, the virus exists in an episomal form, as shown in Figure 4 [19, 45]. Gene expression patterns for hrHPV types change as the host cell reaches different stages of differentiation. Following the initial productive phase, a latent phase can be initiated. During this period, the virus is present in the host, but at very low copy numbers. As a result, the infection is poorly detected by the host immune system [46, 49]. Viral latency can be caused by an infection that did not reach sufficient viral load levels to trigger the immune system during its productive phase. Latency can also be represented by an infection that is detected by the immune system, but subsequently not completely cleared. When the immune pressure subsides, the virus can then reactivate [46, 49, 50].

Worth noting is that viral latency is, at least partially, a semantic concept. An infection could be intermittently positive with a certain genotyping assay, while a more sensitive assay could possibly detect a continuous, persistent infection [50, 51]. As a result, it is difficult to truly discriminate between biologically latent HPV infections and HPV infections that are simply below the level of detection of a given test. Genotyping assays that are clinically validated, such as the GP5+/6+ broad-spectrum PCR, consider infections below the level of detection as clinically irrelevant. The concept and consequences of HPV detection are further discussed in paragraph 2.3.

Cervical carcinogenesis

The path leading up to cervical cancer starts with a persistent HPV infection and is marked by histopathologically distinguishable pre-cancerous states called cervical intraepithelial lesions (CIN). Persistent HPV infections are common and infections with multiple HPV types occur in about 30% of all HPV infections [52, 53]. These complex infections are increased in individuals with impaired immunity, such as HIV patients. However, each HPV type is thought to develop independently into associated CIN lesions [54]. CIN lesions are graded from CIN1 to CIN3 based on the width of dysplastic epithelium compared to normal cervical epithelium. In CIN1, one third of the epithelium is dysplastic, whereas in CIN3, over two thirds of the epithelium is replaced by dysplastic cells. In addition, from CIN1 to CIN3, the dysplastic cells show increasing atypia.

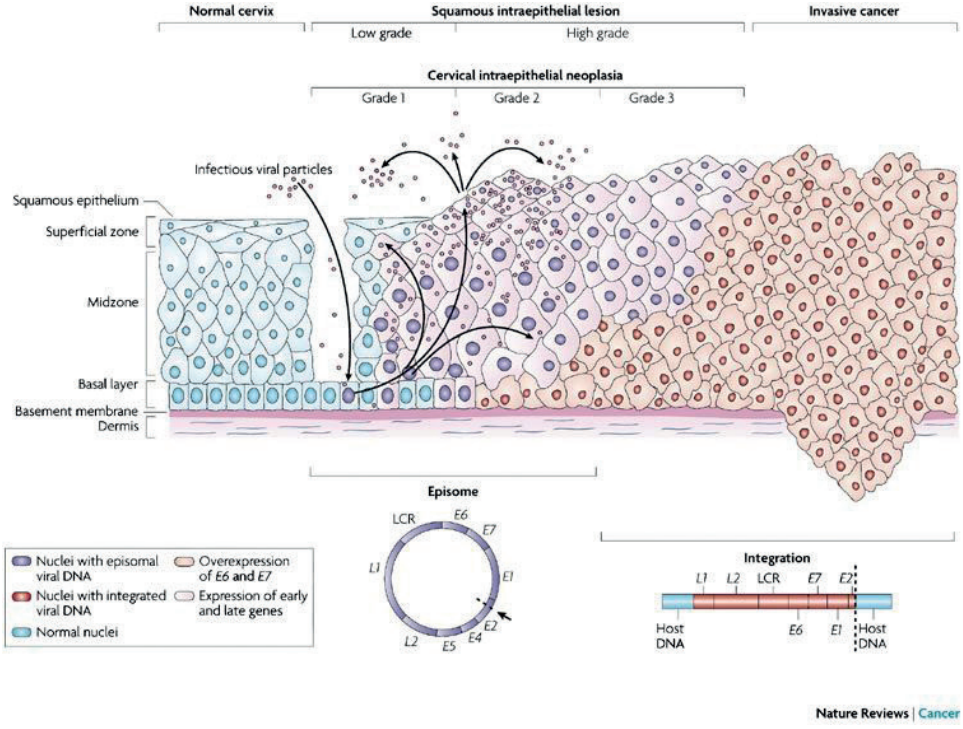


Figure 4: Schematic representation of HPV infection, progression and carcinogenesis. Combined and adapted from Woodman *et al.*, and Steenbergen *et al.* [44, 45]. The different stages of a high-risk HPV infection are displayed, along with the transformations occurring in the cellular tissue.

Persistent HPV infections can be subdivided in productive infections, which are associated with low-grade cervical intraepithelial neoplasia (CIN1/2), and transforming infections, which are associated with high-grade cervical malignancies (CIN2/3), as shown in Figure 4. Due to similar morphology, pathologists are unable to differentiate between productive and transforming CIN2 lesions. The transition from HPV initial infection to CIN2/3 can occur within three to five years [55]. Persistent HPV infections may progress, remain stable or regress and clear. Consequently, CIN lesions of any stage can regress. The likelihood of HPV clearance and thus regression of CIN lesions diminishes with severity of the CIN lesion. On the other hand, progression of HPV infections into invasive cervical cancer is a lengthy process, which can take decades [56]. Since CIN lesions precede the development of cervical cancer, and the progression from CIN to cancer is rather slow, treatment of CIN lesions is highly efficacious in preventing said cancers.

Precursor lesions in non-cervical disease

Precursor lesions have also been described for other HPV related cancers. Anal intraepithelial neoplasia (AIN), penile intraepithelial neoplasia (PIN or PeIN), vulvar intraepithelial neoplasia (VIN) and vaginal intraepithelial neoplasia (VAIN) have been described as relevant precursors for their respective cancers [57–60]. Like in CIN lesions, grades 1–3 indicate the severity of the lesions.

HPV has also been shown to be present in 30–50% of head and neck squamous cell carcinoma (HNSCC) [61, 62]. Interestingly, these HPV positive HNSCC have a different molecular profile and respond much better to therapy compared to HPV negative HNSCC [63, 64]. Unfortunately, to date no defined precursor lesions have been identified for HNSCC making an early detection approach not feasible.

Molecular signature of HPV related cancers

HPV induced cervical cancers display distinct molecular mutation profiles. The two main factors causing these mutations are age and APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) enzymatic activity [65]. Age induces C > T changes in both somatic and germline cells in a NpCpG trinucleotide context, which is presumably caused by spontaneous deamination of cytosine residues. APOBEC is a family of cytidine deaminases and consequentially, APOBEC activity induces C > T changes in a TpCpN context. APOBEC activity also leads to C > G changes, which is postulated to be caused by cytidine deamination, followed by DNA replication and base excision repair errors [65]. Members of the APOBEC family function as innate and possibly adaptive host antiviral responses by hypermutating the HPV genome (and other viral genomes) [66]. On the other hand HPV E6 and E7 have been shown to upregulate APOBEC3A and APOBEC3B activity [67, 68], which leads to an enrichment of APOBEC related changes in HPV induced clinical cancers [65, 67, 69, 70] and potentially drives the generation of HPV diversity [71, 72]. Despite the possible generation of HPV variants by APOBEC activity, it has been shown that conservation of the E7 gene is essential for the development of cervical cancer [43]. While APOBEC activity has been repeatedly implied as an important factor in the development of HPV related disease, it is currently unknown to what extent APOBEC plays a role in early infections.

Detection of HPV infection

Traditionally, screening, detection and monitoring of cervical (pre)cancer was based on cytological results available from (routine) smears. However, in the current era, because of its higher sensitivity for cervical (pre)cancer, HPV testing is becoming increasingly common. Various molecular assays are available to detect and discriminate between hrHPV types in cervical samples. Molecular detection can generally be performed on the DNA, RNA or protein level. RNA and protein detection assays have the advantage of detecting active HPV infections, while DNA detection assays cannot discriminate between active infections and depositions. Since DNA is more stable than RNA and protein, DNA based tests are generally used. Most commonly and depending on the research setting, a broad-spectrum PCR is performed to target hrHPV genotypes. Broad-spectrum

PCRs target conserved regions of the HPV genome, generally L1 or E7 are used. Between the conserved primer sites are regions which allow for discrimination between HPV types. These regions can be targeted to detect different HPV types. This is most commonly performed by using qPCR probes or reverse line blot assays.

Sensitivity of HPV testing and intended use

Molecular HPV detection, by means of broad-spectrum or type specific PCR, has proven its value in various sensitivity and specificity comparisons with traditional cytology [73–75]. Each assay has specific properties with regard to the HPV types being detected. These properties follow an intended use that was defined during assay development. The intended use of a test defines the context in which each test is appropriate for use. For HPV detection, the most common approaches are epidemiological or clinical in nature. Tests that are intended for use in epidemiological contexts, such as HPV surveillance in vaccine efficacy and monitoring studies, require high sensitivity and a broad detection spectrum of HPV types. Tests intended for use in a clinical setting, for example as part of screening programmes or triage strategies, should detect clinically relevant infections. At the same time, these tests should exclude clinically irrelevant infections [51]. Given the specific nature of clinically relevant HPV testing, guidelines have been established for clinical test validation. These guidelines describe criteria, which tests should meet to be adequate for this intended use [76]. These guidelines have resulted in a comprehensive assessment of available tests, resulting in a framework protocol for the validation of current and future tests [77, 78].

For example, the MY09/11 PCR [79] and its more sensitive derivative PGMY [80] form the basis of the Roche Linear Array (LA) genotyping platform. The LA test is intended for use on clinical specimens and can detect 37 different HPV genotypes, although it is clinically validated only for the detection of hrHPV types [81]. Another commonly used clinically validated test makes use of the GP5+/6+ PCR followed by reverse line-blotting for the detection of hrHPV genotypes [82, 83]. For primary screening purposes, the Roche Cobas 4800 HPV detection test was clinically validated [84]. This test generates a more minimalistic output, by only genotyping HPV16 and HPV18 and giving a pooled result for hrHPV positivity. The Cobas 4800 platform is specifically designed for high-throughput capacity by automating sample preparation, making it well-suited for large sample flows generated by screening programmes.

The most sensitive PCR for the detection of HPV infections is the SPF10, which is part of the SPF10-DEIA-LiPA$_{25}$ platform [85, 86]. This platform detects 25 HPV genotypes, including all hrHPV types at very high sensitivity. Due to its detection of clinically irrelevant infections, this test is not clinically validated. Since this thesis primarily focuses on the monitoring of vaccine effects, which implies even very low viral copy number infections should be detected, the SPF10-DEIA-LiPA$_{25}$ platform was the test of choice for genotyping results.

Quantitative detection of HPV infection

The detection of HPV DNA via (broad-spectrum) PCR results in an absolute outcome of HPV types being present or not. However, the amount of viral DNA, or viral load, present in infections is a continuous variable. Viral load has been implicated as an indicator for infection persistence and clearance [87, 88] and may be used as a discriminatory factor in specific cases to rule out CIN3 for women with normal cytology [89]. Viral load is accurately measured through quantitative PCR (qPCR) assays, which rely on an increasing fluorescent signal during amplification to measure the amount of HPV DNA present in a sample. The most common approach is through specific probes which consist of a fluorescent label paired to an oligonucleotide which will bind to a complementary DNA sequence. To ensure the fluorescent label does not produce any background signal, probes often contain quenchers as well. During replication with a polymerase that contains 5'-3' exonuclease activity, the probe will be degraded, resulting in increased fluorescence signal, which can be measured over time.

Viral load quantification can be performed in multiplex, with similar sensitivity limitations as for conventional broad-spectrum PCRs. However, additionally, the viral load itself could be affected by competition for reagents when multiple HPV types are present in the same sample. Therefore, in specific cases, type-specific quantifications can be preferred, especially for low viral load infections. This is the case in vaccine monitoring studies, where maximum sensitivity and reliability are required, as described in this thesis.

Advantages and disadvantages of multiplex HPV detection

Simultaneous detection of multiple HPV genotypes in a single test is extremely efficient with regard to time, money and sample materials. Therefore, it is the preferred method of choice in many epidemiological and vaccine monitoring studies [6, 90–94]. Multiplex HPV genotyping allows for the assessment of type replacement effects following vaccination. *Type replacement* occurs when another (similar) HPV type occupies the niche that has become available by the vaccine reducing prevalence of another HPV type, which has been shown to occur following pneumococcal vaccination [95, 96]. In addition, type replacement has been proposed to be theoretically possible for HPV [97]. To get a broad overview of possible type replacement effects, most multiplex HPV genotyping assays include at least all group 1 carcinogenic HPV types. Type replacement is continuously being investigated in vaccine monitoring studies, and so far, no definitive negative vaccine effects have been observed [98], although there are some caveats.

For instance, multiplex testing introduces competition for reagents when multiple HPV types are present in a sample, which is common for HPV [52, 53]. This could result in the underdetection of HPV types in the presence of specific other HPV types. Following vaccination, certain HPV types could become less prevalent, possibly leading to the increased detection of other HPV types, which were previously not detected due to competition within the multiplex assay. This process is called unmasking and can occur in all multiplex assays where competition for reagents occur, although the exact competition between types differs per assay [51, 99].

Results from studies that monitor an intervention using multiplex based assays should always be considered in the light of unmasking, before proper type replacement effects can be established.

When interpreting results, one should be thoroughly aware of which multiplex test was used and how that test could affect the study outcome. Since for each multiplex test, the pattern of competition is different, method studies should be conducted to generate proper test-specific profiles. For example, HPV52 has been reported more frequently following vaccination in some studies [100, 101]. On the other hand, it has been suggested that unmasking causes enhanced detection of HPV52 in cohorts tested with PGMY based genotyping assays [102]. Worth noting, is that despite increased costs in labour, expenses and study materials, type-specific detection in singleplex assays circumvents the issue of unmasking entirely. A possible solution to rule out possible effects of unmasking when type replacement is potentially observed, would be to retest a cohort using type-specific assays.

HPV detection by sequencing: Sanger sequencing

Genotyping of HPV infections allows for the specific identification of genotypes to study various hypotheses. However, to establish viral evolution and specific sequence effects in the context of infection and disease development, insight into the viral sequence is required. The golden standard for sequencing is Sanger or capillary sequencing. This method generates relatively long reads (> 1000 basepairs) which, if they overlap, can be assembled to generate complete genomes. However, Sanger sequencing generates low resolution results, meaning that the amount of times a specific nucleotide is covered by reads is low. This implies that if different nucleotides occur at a single position, generally only the most dominant one can be reported. Other disadvantages of Sanger sequencing are the low throughput of the method and the dependency on amplified sample materials. Despite its disadvantages, Sanger sequencing remains a potent tool to answer questions which only require knowledge of the dominant variant involved in an HPV infection.

HPV detection by sequencing: Next-generation sequencing

Currently, next-generation sequencing (NGS) is a strong alternative for Sanger sequencing. NGS methods generate exponentially more data, potentially resulting in extremely high sequence coverage. High coverage allows for stochastic determination of nucleotide ratios across the genome. NGS is highly scalable, potentially allows for agnostic (PCR-free) detection of sequences, and generates reads of differing lengths depending on the NGS method used. Regardless of read length, the overlap between reads is important to generate an assembly of complete genomes. NGS opens up possibilities for population-wide HPV screening studies and in-depth assessment of HPV evolution. Currently, the clinical value of NGS is under investigation.

Prevention of cervical cancer

Cervical cancer is a preventable disease marked by a persistent HPV infection preceding histopathologically clearly defined precursor lesions. Three types of prevention have been defined.

- Primary prevention deals with prevention of disease in healthy people.
- Secondary prevention aims to detect and treat asymptomatic disease, preventing progression to symptomatic disease.
- Tertiary prevention is aimed at burden reduction of symptomatic disease.

Primary prevention of cervical cancer occurs through prophylactic vaccination against HPV infection. Secondary prevention of cervical cancer occurs via (molecular) screening programmes. Finally, tertiary prevention consists of treatment of CIN2/3 lesions. Following treatment, recurrence of HPV related disease is relatively common. As a result, post-treatment follow-up of CIN2/3 is advised, which can also be considered a form of tertiary prevention. Since this thesis focuses on the primary prevention of HPV related disease, secondary and tertiary prevention will only be summarily discussed.

Prophylactic HPV vaccination

The discovery of a causal relationship between HPV and cervical (and later other forms of) cancer by Harald zur Hausen in [103], has led to an inquest into targeting strategies against the virus and eventually resulted in the development of highly efficient prophylactic HPV vaccines [104]. A major breakthrough towards vaccine development was the discovery that the L1 protein was able to form immunogenic virus-like particles through self-assembly [105–107].

Vaccine properties

At the time of writing, three prophylactic HPV vaccines are licensed in Europe and commercially available. The bivalent vaccine (Cervarix, developed by GSK), targets HPV16 and HPV18, thereby potentially preventing around 70% of worldwide cervical cancer cases [6]. The quadrivalent vaccine (Gardasil4, developed by Merck), besides protection against HPV16 and HPV18, adds protection against HPV6 and HPV11, thereby potentially preventing up to 90% of genital warts cases [108]. More recently, the nonavalent vaccine (Gardasil 9, also developed by Merck) was registered in Europe. This vaccine prevents against the same types as the quadrivalent vaccine, but adds HPV31, 33, 45, 52 and 58, thereby potentially preventing up to 90% of all cervical cancers worldwide [6]. The main differences between the vaccines produced by Merck and the one produced by GSK, are in the dosage of the virus-like particles, the use of different expression systems and in the use of different adjuvants, as shown in Table 3 [109]. The bivalent and quadrivalent

vaccines were originally registered to protect against cervical cancer. Licensure has since then been extended to protection against non-cervical HPV associated disease, and for use in boys. All three available vaccines make use of L1 virus-like particles (VLP) and are highly efficacious against the vaccine HPV types [90–92, 109–117].

It should be noted that in reporting (vaccine) efficacies, clinical trials differentiate between according-to-protocol populations and intention-to-treat populations. For HPV, the according-to-protocol analysis is performed on seronegative and HPV DNA negative study participants. The intention to treat analysis is performed on all participants enrolled in a study, regardless of prior HPV status, thereby approaching a real-life situation.

Vaccination primarily leads to antibody-mediated sterilizing immunity against HPV infection. Upon vaccination, high serum titres of VLP induced antibodies are established. *In vitro* assays have demonstrated type-specific HPV neutralization of these antibodies. The antibodies prevent binding of free virus to the basement membrane, thereby preventing infection [118]. Since this mechanism functions by preventing HPV infection, it is important that the vaccine be given to HPV-naïve or HPV DNA negative recipients. Vaccine trials have shown that vaccine efficacy is strongly affected by increasing time since first sexual intercourse and by an increasing number of sexual partners [119]. These findings indicate that with an increased risk of having attained an HPV infection prior to vaccination, the vaccine efficacy is reduced. Such effects become apparent during analysis of the intention to treat population of clinical trials. Besides the antibody-mediated effects, a T cell-mediated immune response is also generated upon vaccination [118, 120]. The mechanism of the HPV vaccine induced T-cell response is not yet fully understood, but it is thought that cross-protective effects of the vaccines against non-vaccine types might be the result of this T-cell response [109, 118].

Efficacy measured via intermediary endpoints

Development from initial HPV infection to cervical cancer takes decades [56]. Therefore, using cancer as a disease end-point is impractical and unethical for clinical trials assessing vaccine efficacy. As a result, the value of intermediate endpoints was assessed for determining vaccine efficacy against cervical cancer in clinical trials. Since CIN lesions are the precursor lesions for cervical cancer, the most advanced CIN lesions (CIN3 and later CIN2) were approved as disease end-points for the phase III clinical trials for Cervarix and Gardasil4 [90, 110, 115]. Although CIN2/3 develop considerably faster than cervical cancer (2.5–4 years), the time between vaccination and disease end-point is still substantial. As a result, virological end-points (i.e. persistent HPV infections) were introduced as surrogate end-points. In 2014, the HPV working group from the International Agency for Cancer Research advised that persistent HPV infections 6 months or longer were viable intermediate end-points for the evaluation of vaccine protection. However, it should be noted that persistent infections cannot replace actual disease endpoints and should therefore only be used in appropriate settings. In vaccine efficacy studies, protection has to be assessed using >6-month persistent infections, but disease end-points have to be verified by monitoring post-licensure [122].

Table 3: Compositions of the three available HPV vaccines, adapted from [121].

Vaccine	Composition	Adjuvant
Cervarix (bivalent)	20 µg HPV16 L1 protein	AS04
	20 µg HPV18 L1 protein	
Gardasil (quadrivalent)	20 µg HPV6 L1 protein	AAHS
	40 µg HPV11 L1 protein	
	40 µg HPV16 L1 protein	
	20 µg HPV18 L1 protein	
Gardasil9 (nonavalent)	30 µg HPV6 L1 protein	AAHS
	40 µg HPV11 L1 protein	
	60 µg HPV16 L1 protein	
	40 µg HPV18 L1 protein	
	20 µg HPV31 L1 protein	
	20 µg HPV33 L1 protein	
	20 µg HPV45 L1 protein	
	20 µg HPV52 L1 protein	
	20 µg HPV58 L1 protein	

AS04: Adjuvant system 04 (aluminum hydroxide and monophosphoryl lipid A)
AAHS: Amorphous aluminum hydroxyphosphate sulfate adjuvant

Cross-protection against non-vaccine HPV types

Besides excellent protection of the bivalent and quadrivalent vaccine against their respective targeted HPV types, both vaccines have also shown cross-protective effects against non-vaccine oncogenic HPV types [91, 109, 114, 117, 123, 124]. The previously mentioned added value of disease endpoints becomes apparent in Table 4. A considerably higher VE is found against CIN3/CIN2, than against persistent infection [109]. The bivalent vaccine seems also proficient in preventing high-grade cervical disease caused by high-risk non-vaccine HPV types. It has been suggested that the bivalent vaccine has a similar efficacy against CIN2+ as the nonavalent vaccine, as shown in Table 4 [109]. In addition to (limited) protection against oncogenic HPV types, the bivalent vaccine has also been suggested to have some effect against HPV6 and HPV11 infections [125, 126], although this could not be confirmed in a cross-sectional observational study in the Netherlands [127]. The strong cross-protective effects of the bivalent vaccine are thought to be caused by the use of the AS04 adjuvant. This adjuvant elicits higher antibody titers, which persist longer than those reported for aluminum hydroxide salt adjuvant containing HPV vaccines only (Gardasil) [128]. In addition, a higher B cell memory response is generated by the bivalent vaccine [128].

Vaccine cross-protection appears to be limited to HPV types phylogenetically related to the vaccine types [91, 109, 114, 117, 123, 124]. Despite cross-protective effects being higher against CIN2+ than against 6-month persistent infections, cross-protection is weaker than protection against vaccine types. This led to the hypothesis that cross-protection might be explained by protection against specific lineages of non-vaccine HPV types, rather than all variants. The amount of studies considering lineage specific vaccine cross-protective effects are limited. However, a study by Harari *et al.* investigating this question found that vaccine effectiveness against HPV16, 18, 33, 35, 45 and 51 was not significantly different for any of the lineages of these types [129]. This implies that for types against which partial cross-protection occurs, the level of cross-protection is not explained by the lineages occurring for that HPV type, although more studies into vaccine effects against HPV type lineages are required.

Table 4: Vaccine efficacies of the three registered HPV vaccines against both vaccine and non-vaccine HPV types in women who were both seronegative and HPV DNA negative for each tested HPV type (according-to-protocol). Adapted from Harper *et al.* [109].

	Gardasil	Gardasil9	Cervarix
Among women 15/16–26 years			
4–6 months HPV 16/18 infection	96% (83–100)	na	94% (92–96)
6 month HPV 31/33/45/52/58 infection	18% (5–29)	96% (94–98)	na
6 month HPV 31 infection	46% (15–66)	96% (91–98)	77% (69–83)
6 month HPV 33 infection	NS	99% (95–100)	45% (25–60)
6 month HPV 45 infection	NS	97% (92–99)	74% (58–84)
6 month HPV 51 infection	na	na	17% (4–28)
6 month HPV 52 infection	NS	97% (95–99)	na
6 month HPV 58 infection	NS	95% (91–97)	na
CIN 2 + related to HPV 16/18	98% (94–100)	na	98% (88–100)
CIN 2 + related to HPV 31	70% (32–88)	100% (40–100)	88% (68–96)
CIN 2 + related to HPV 33	NS	100% (33–100)	68% (40–84)
CIN 2 + related to HPV 39	NS	na	75% (22–94)
CIN 2 + related to HPV 45	NS	NS	82% (17–98)
CIN 2 + related to HPV 51	NS	na	54% (22–74)
CIN 2 + related to HPV 52	NS	100% (67–100)	na
CIN 2 + related to HPV 58	NS	NS	na
CIN 2 + caused by any HPV type	**22% (3–38)**	**63% (35–79)**	**62% (47–73)**
CIN 3 + caused by any HPV type	**43% (24–57)**	**na**	**93% (79–99)**
AIS caused by any HPV type	na	na	100% (31–100)

Vaccine efficacies are presented with 95% confidence intervals.
NS = not significant; na = not applicable/available.
Bold signifies the clinically important outcomes.

Vaccine implementation in the Netherlands

Bivalent HPV vaccination was implemented in the Dutch National Immunization Program (NIP) in 2009. At the time, a three-dose schedule at 0, 1 and 6 months was followed. Vaccination, which is voluntary in the Netherlands, started with a catch-up campaign for girls born between 1993 and 1996. From 2010 and onwards, girls are invited for vaccination in the year they turn thirteen. In the following years, and continuing this day, the three-dose schedule vaccination was monitored in the Netherlands by the National Institute for Public Health and the Environment (RIVM) [93, 94, 130, 131]. In 2011 and 2014 it was shown that bivalent vaccination according to a two-dose schedule (0 and 6 months) was not immunologically inferior to the three-dose schedule [132–134]. As a result, following a new recommendation in 2014 by the European Medicines Agency, the HPV vaccination program in the Netherlands transitioned to a two-dose schedule at 0 and 6 months [135]. This thesis will primarily focus on monitoring the three-dose schedule, as results from the two-dose schedule monitoring study are not yet available.

Secondary prevention of cervical cancer in the Netherlands

Secondary prevention is the early detection and treatment of disease in patients with subclinical disease. Screening for cervical cancer by cytology was introduced regionally in the 1970's and was spread across the country in subsequent years. Since 1996, cervical screening is organized as a nationwide programme, with cytology as a primary test. This screening programme resulted in a strong reduction of the incidence and mortality of cervical cancer [136]. Cytology as a screening tool was successful despite its limited sensitivity for cervical cancer (60–65% for CIN3+), due to the frequent repeat testing employed in the screening programme [137, 138]. Given the long lead time to cervical cancer (15–30 years), cytology missing some abnormal smears was permitted. Recently however, no further decrease in cervical cancer incidence has been observed, suggesting that the maximum yield of cytology-based screening was reached [136, 139]. To further reduce cervical cancer incidence, HPV testing for cervical screening was considered. HPV testing was convincingly shown to be more sensitive at CIN3+ detection than cytology, thus providing better protection against cervical cancer and CIN3 [73–75, 139, 140]. This development has led to the new HPV based screening programme being implemented in the Netherlands in 2017 [141].

The new screening programme introduced ten-year screening intervals for women aged 40–50 years, who have a negative HPV test. In addition, women who do not respond to the screening invitation, are now offered self-collection kits. Since HPV testing has a lower specificity for cervical cancer than cytology, a triage test for colposcopy referral was introduced to identify HPV positive women at risk for disease. Cytology is used as a triage test at baseline, with additional cytology at 6 months. Women are referred for colposcopy when they have an abnormal smear (threshold ≥ASC-US or PAP2/3A1). HPV positive women with negative cytology tests at baseline and 6 months are sent back to cervical screening at a 5-year interval even when they are between 40 and 50 years, because these women have a significantly elevated 5-year CIN3+ risk [73]. At age

60, women exit the screening programme if negative by HPV testing. In case of a positive HPV test with negative repeat cytology, an additional screening round is performed at age 65. The first results of the first year of screening programme have recently been published [141].

Table 5: Long-term cancer, CIN3+ and CIN2+ risks in HPV-negative women (A) and HPV-positive women with negative triage (B). Adapted from [139].

Cohort	Follow-up period	Cancer risk	CIN3+ risk	CIN2+ risk
A. HPV-negative women				
VUSA-Screen[a]	5 years	-	0.09%	0.21%
POBASCAM[b]	14 years	0.09%	0.56%	-
ARTISTIC[c]	6 years	-	0.28%	0.87%
Swedescreen[d]	13 years	-	0.84%	1.74%
Kaiser Permanente[e]	18 years	-	0.90%	1.85%
B. HPV-positive, triage negative women				
Triage with cytology and repeat cytology				
VUSA-Screen[a]	5 years	-	4.1%	7.0%
POBASCAM[b]	14 years	-	±10.4%	-
Triage with cytology and HPV16/18 genotyping				
VUSA-Screen[a]	5 years	-	3.5%	7.9%
POBASCAM[b]	14 years	-	±8.7%	-

CIN3+ = cervical intraepithelial neoplasia grade 3 or worse
CIN2+ = cervical intraepithelial neoplasia grade 2 or worse
[a]: Uijterwaal *et al.*, Cancer Prev. Res., 2015
[b]: Dijkstra *et al.*, BMJ, 2016
[c]: Kitchener *et al.*, Eur. J. Cancer, 2011
[d]: Elfstrom *et al.*, BMJ, 2014
[e]: Castle *et al.*, J. Clin. Oncol., 2012

Post-treatment monitoring of women with CIN2/3

Recurrence rates of CIN2/3 after treatment by LLETZ (Large Loop Excision of the Transformation Zone) or LEEP (Loop Electrosurgical Excision Procedure) are 5–15% within two years [142, 143]. As a result, in the Netherlands, women treated for CIN2/3 are monitored either by cytology alone or by a combination of cytology and molecular HPV detection (cotesting) 6, 12 and 24 months after treatment [143, 144]. If during contesting the 6-month visit shows both a negative HPV test and normal cytology, the 12-month visit can be omitted. If three consecutive visits show normal cytology, or the 6- and 24-month visits have negative HPV and cytology co-tests, the woman is referred back to the population-based screening programme [143, 145]. Recurrent CIN can be caused by incomplete removal of the lesion, or by a newly acquired HPV infection with either a new HPV type or a variant of the same HPV genotype.

Aim of this thesis

This thesis explores the value and relevance of viral load detection and HPV variant analysis in monitoring of HPV infections in vaccinated and non-vaccinated women. I will discuss epidemiological vaccine monitoring studies and clinical studies, starting with pre-vaccination baseline studies to identify the potential use of the developed molecular techniques.

In **part I**, the focus is on viral load as a marker for clearing and persistent infections.

In **chapter 2**, we developed HPV16 and HPV18 specific qPCR assays to analyze viral load for these types. Population level trends are assessed and compared to previous literature. In addition, the value of viral load testing in predicting persisting infections at the individual level is described.

Chapter 3 continues this train of thought, by development of qPCR assays against all other high-risk HPV types (n = 11), as well as low-risk HPV6 and HPV11, which cause > 90% of all genital warts. We applied the developed assays to a cohort of young women who were either non-vaccinated or fully vaccinated (3 doses), to study possible vaccine effects on type-specific viral loads.

In order to evaluate the potential effect of the bivalent vaccine on the prevalence of HPV variants we set out in **part II** to identify viral diversity of HPV16 and HPV18 in non-vaccinated women.

In **chapter 4**, we developed a whole genome sequencing assay for HPV16 to study and describe the molecular diversity of HPV16 in a cohort of young, non-vaccinated women. Within this study both clearing and persistent infections are sequenced and compared for possible differences.

In **chapter 5**, we investigate the viral diversity of HPV18, for which a whole genome sequencing assay was developed as well. As available sequence data for HPV18 was especially scarce, a phylogenetic comparison was performed, comparing Dutch sequencing data to the Genbank database.

In **chapter 6**, we shift from classical Sanger sequencing to high-resolution Illumina sequencing to study the occurrence of minority variants in a subset of HPV16 samples, from the same cohort of young non-vaccinated women as described in chapter 4. Special attention was given to host APOBEC mediated genome editing, as the literature describes a role for these enzymes in cervical carcinogenesis. We also paid attention to other potential mutation signatures in infections in young women. The identified mutation signatures were compared to those described in previous studies. Furthermore, the mutation signatures identified from different samples from the same study participant were compared to assess trends over time.

In **chapter 7**, we adapted the sequencing assay described in Chapter 4 for high-resolution NGS. The adapted assay was used to investigate whether recurrent CIN after CIN2/3 treatment is caused by incomplete removal of the lesion, or by a newly acquired HPV infection with either a new HPV type, or a variant of the same HPV genotype. We compared paired CIN2/3 samples prior to treatment with 6–12-month recurrent lesions.

In **chapter 8**, we investigate the association between the degree of cross-protection against non-vaccine HPV types and the phylogenetic distance between vaccine and non-vaccine HPV types. This assessment was performed using cross-sectional data from a large study with both vaccinated and non-vaccinated women. Phylogenetic distance was estimated based on reference sequences for all HPV types included in the SPF_{10}-DEIA-LiPA25 platform.

References

1. Baseman, J.G. and L.A. Koutsky, *The epidemiology of human papillomavirus infections.* J Clin Virol, 2005. **32 Suppl 1**: p. S16–24.
2. de Martel, C., et al., *Worldwide burden of cancer attributable to HPV by site, country and HPV type.* Int J Cancer, 2017. **141**(4): p. 664–670.
3. Walboomers, J.M., et al., *Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.* J Pathol, 1999. **189**(1): p. 12–9.
4. Gissmann, L., et al., *Presence of human papillomavirus in genital tumors.* J Invest Dermatol, 1984. **83**(1 Suppl): p. 26s-28s.
5. de Sanjose, S., et al., *Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis.* Lancet Infect Dis, 2007. **7**(7): p. 453–9.
6. de Sanjose, S., et al., *Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study.* Lancet Oncol, 2010. **11**(11): p. 1048–56.
7. Bernard, H.U., et al., *Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments.* Virology, 2010. **401**(1): p. 70–9.
8. de Villiers, E.M., *Cross-roads in the classification of papillomaviruses.* Virology, 2013. **445**(1–2): p. 2–10.
9. Bzhalava, D., C. Eklund, and J. Dillner, *International standardization and classification of human papillomavirus types.* Virology, 2015. **476**: p. 341–4.
10. Humans, I.W.G.o.t.E.o.C.R.t., *Biological agents. Volume 100 B. A review of human carcinogens.* IARC Monogr Eval Carcinog Risks Hum, 2012. **100**(Pt B): p. 1–441.
11. Egawa, N. and J. Doorbar, *The low-risk papillomaviruses.* Virus Res, 2017. **231**: p. 119–127.
12. de Sanjose, S., M. Brotons, and M.A. Pavon, *The natural history of human papillomavirus infection.* Best Pract Res Clin Obstet Gynaecol, 2018. **47**: p. 2–13.
13. Bergvall, M., T. Melendy, and J. Archambault, *The E1 proteins.* Virology, 2013. **445**(1–2): p. 35–56.
14. McBride, A.A., *The papillomavirus E2 proteins.* Virology, 2013. **445**(1–2): p. 57–79.
15. Hoppe-Seyler, K., et al., *The HPV E6/E7 Oncogenes: Key Factors for Viral Carcinogenesis and Therapeutic Targets.* Trends Microbiol, 2018. **26**(2): p. 158–168.
16. Roman, A. and K. Munger, *The papillomavirus E7 proteins.* Virology, 2013. **445**(1–2): p. 138–68.
17. Vande Pol, S.B. and A.J. Klingelhutz, *Papillomavirus E6 oncoproteins.* Virology, 2013. **445**(1–2): p. 115–37.
18. Steinbach, A. and A.B. Riemer, *Immune evasion mechanisms of human papillomavirus: An update.* Int J Cancer, 2018. **142**(2): p. 224–229.
19. Moody, C., *Mechanisms by which HPV Induces a Replication Competent Environment in Differentiating Keratinocytes.* Viruses, 2017. **9**(9).
20. Christiansen, I.K., et al., *Transcriptionally active regions are the preferred targets for chromosomal HPV integration in cervical carcinogenesis.* PLoS One, 2015. **10**(3): p. e0119566.
21. Kim, K., et al., *Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection.* J Virol, 2003. **77**(23): p. 12450–9.
22. Kajitani, N., et al., *Productive Lifecycle of Human Papillomaviruses that Depends Upon Squamous Epithelial Differentiation.* Front Microbiol, 2012. **3**: p. 152.
23. Van Doorslaer, K., *Evolution of the papillomaviridae.* Virology, 2013. **445**(1–2): p. 11–20.
24. Rector, A. and M. Van Ranst, *Animal papillomaviruses.* Virology, 2013. **445**(1–2): p. 213–23.

25. Bernard, H.U., *Coevolution of papillomaviruses with human populations.* Trends Microbiol, 1994. **2**(4): p. 140–3.
26. Doorbar, J., et al., *Human papillomavirus molecular biology and disease association.* Rev Med Virol, 2015. **25 Suppl 1**: p. 2–23.
27. de Villiers, E.M., et al., *Classification of papillomaviruses.* Virology, 2004. **324**(1): p. 17–27.
28. Burk, R.D., A. Harari, and Z. Chen, *Human papillomavirus genome variants.* Virology, 2013. **445**(1–2): p. 232–43.
29. Chen, Z., et al., *Classification and evolution of human papillomavirus genome variants: Alpha-5 (HPV26, 51, 69, 82), Alpha-6 (HPV30, 53, 56, 66), Alpha-11 (HPV34, 73), Alpha-13 (HPV54) and Alpha-3 (HPV61).* Virology, 2018. **516**: p. 86–101.
30. Chen, Z., et al., *Evolutionary dynamics of variant genomes of human papillomavirus types 18, 45, and 97.* J Virol, 2009. **83**(3): p. 1443–55.
31. Chen, Z., et al., *Evolution and taxonomic classification of alphapapillomavirus 7 complete genomes: HPV18, HPV39, HPV45, HPV59, HPV68 and HPV70.* PLoS One, 2013. **8**(8): p. e72565.
32. Ho, L., et al., *The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations.* J Virol, 1993. **67**(11): p. 6413–23.
33. Yamada, T., et al., *Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective.* J Virol, 1997. **71**(3): p. 2463–72.
34. Clifford, G.M., et al., *Human Papillomavirus 16 sub-lineage dispersal and cervical cancer risk worldwide: whole viral genome sequences from 7116 HPV16-positive women.* Papillomavirus Res, 2019.
35. Xi, L.F., et al., *Human papillomavirus type 16 and 18 variants: race-related distribution and persistence.* J Natl Cancer Inst, 2006. **98**(15): p. 1045–52.
36. Andersson, S., et al., *Uneven distribution of HPV 16 E6 prototype and variant (L83V) oncoprotein in cervical neoplastic lesions.* Br J Cancer, 2000. **83**(3): p. 307–10.
37. Lizano, M., et al., *Distribution of HPV16 and 18 intratypic variants in normal cytology, intraepithelial lesions, and cervical cancer in a Mexican population.* Gynecol Oncol, 2006. **102**(2): p. 230–5.
38. Xi, L.F., et al., *Risk for high-grade cervical intraepithelial neoplasia associated with variants of human papillomavirus types 16 and 18.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(1): p. 4–10.
39. Sichero, L., J.S. Sobrinho, and L.L. Villa, *Oncogenic potential diverge among human papillomavirus type 16 natural variants.* Virology, 2012. **432**(1): p. 127–32.
40. Zhang, L., et al., *Variants of human papillomavirus type 16 predispose toward persistent infection.* Int J Clin Exp Pathol, 2015. **8**(7): p. 8453–9.
41. Cullen, M., et al., *Deep sequencing of HPV16 genomes: A new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection.* Papillomavirus Research, 2015.
42. Mirabello, L., et al., *HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women.* J Natl Cancer Inst, 2016. **108**(9).
43. Mirabello, L., et al., *HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis.* Cell, 2017. **170**(6): p. 1164–1174 e6.
44. Woodman, C.B., S.I. Collins, and L.S. Young, *The natural history of cervical HPV infection: unresolved issues.* Nat Rev Cancer, 2007. **7**(1): p. 11–22.
45. Steenbergen, R.D., et al., *Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions.* Nat Rev Cancer, 2014. **14**(6): p. 395–405.

46. Doorbar, J., *The papillomavirus life cycle.* J Clin Virol, 2005. **32 Suppl 1**: p. S7–15.
47. Rositch, A.F., et al., *Patterns of persistent genital human papillomavirus infection among women worldwide: a literature review and meta-analysis.* Int J Cancer, 2013. **133**(6): p. 1271–85.
48. Schiffman, M., et al., *Human papillomavirus and cervical cancer.* Lancet, 2007. **370**(9590): p. 890–907.
49. Doorbar, J., *Latent papillomavirus infections and their regulation.* Curr Opin Virol, 2013. **3**(4): p. 416–21.
50. Doorbar, J., *Host control of human papillomavirus infection and disease.* Best Pract Res Clin Obstet Gynaecol, 2018. **47**: p. 27–41.
51. Snijders, P.J., A.J. van den Brule, and C.J. Meijer, *The clinical relevance of human papillomavirus testing: relationship between analytical and clinical sensitivity.* J Pathol, 2003. **201**(1): p. 1–6.
52. Dickson, E.L., et al., *Multiple-type human papillomavirus (HPV) infections: a cross-sectional analysis of the prevalence of specific types in 309,000 women referred for HPV testing at the time of cervical cytology.* Int J Gynecol Cancer, 2013. **23**(7): p. 1295–302.
53. Schmitt, M., et al., *Multiple human papillomavirus infections with high viral loads are associated with cervical lesions but do not differentiate grades of cervical abnormalities.* J Clin Microbiol, 2013. **51**(5): p. 1458–64.
54. Quint, W., et al., *One virus, one lesion—individual components of CIN lesions contain a specific HPV type.* J Pathol, 2012. **227**(1): p. 62–71.
55. Snijders, P.J., et al., *HPV-mediated cervical carcinogenesis: concepts and clinical implications.* J Pathol, 2006. **208**(2): p. 152–64.
56. Vink, M.A., et al., *Clinical progression of high-grade cervical intraepithelial neoplasia: estimating the time to preclinical cervical cancer from doubly censored national registry data.* Am J Epidemiol, 2013. **178**(7): p. 1161–9.
57. Roberts, J.R., L.L. Siekas, and A.M. Kaz, *Anal intraepithelial neoplasia: A review of diagnosis and management.* World J Gastrointest Oncol, 2017. **9**(2): p. 50–61.
58. Preti, M., et al., *Vulvar intraepithelial neoplasia.* Best Pract Res Clin Obstet Gynaecol, 2014. **28**(7): p. 1051–62.
59. Demeter, L.M., et al., *Penile intraepithelial neoplasia: clinical presentation and an analysis of the physical state of human papillomavirus DNA.* J Infect Dis, 1993. **168**(1): p. 38–46.
60. Sillman, F.H., et al., *Vaginal intraepithelial neoplasia: risk factors for persistence, recurrence, and invasion and its management.* Am J Obstet Gynecol, 1997. **176**(1 Pt 1): p. 93–9.
61. Gillison, M.L. and D.R. Lowy, *A causal role for human papillomavirus in head and neck cancer.* Lancet, 2004. **363**(9420): p. 1488–9.
62. Kimple, R.J. and P.M. Harari, *The prognostic value of HPV in head and neck cancer patients undergoing postoperative chemoradiotherapy.* Ann Transl Med, 2015. **3**(Suppl 1): p. S14.
63. Ang, K.K., et al., *Human papillomavirus and survival of patients with oropharyngeal cancer.* N Engl J Med, 2010. **363**(1): p. 24–35.
64. Fakhry, C., et al., *Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial.* J Natl Cancer Inst, 2008. **100**(4): p. 261–9.
65. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415–21.
66. Stavrou, S. and S.R. Ross, *APOBEC3 Proteins in Viral Immunity.* J Immunol, 2015. **195**(10): p. 4565–70.

67. Warren, C.J., et al., *APOBEC3A functions as a restriction factor of human papillomavirus.* J Virol, 2015. **89**(1): p. 688–702.
68. Mori, S., et al., *Human Papillomavirus 16 E6 Upregulates APOBEC3B via the TEAD Transcription Factor.* J Virol, 2017. **91**(6).
69. Burns, M.B., N.A. Temiz, and R.S. Harris, *Evidence for APOBEC3B mutagenesis in multiple human cancers.* Nat Genet, 2013. **45**(9): p. 977–83.
70. Burns, M.B., et al., *APOBEC3B is an enzymatic source of mutation in breast cancer.* Nature, 2013. **494**(7437): p. 366–70.
71. Warren, C.J., et al., *Roles of APOBEC3A and APOBEC3B in Human Papillomavirus Infection and Disease Progression.* Viruses, 2017. **9**(8).
72. Hirose, Y., et al., *Within-Host Variations of Human Papillomavirus Reveal APOBEC Signature Mutagenesis in the Viral Genome.* J Virol, 2018. **92**(12).
73. Rijkaart, D.C., et al., *Human papillomavirus testing for the detection of high-grade cervical intraepithelial neoplasia and cancer: final results of the POBASCAM randomised controlled trial.* Lancet Oncol, 2012. **13**(1): p. 78–88.
74. Ogilvie, G.S., et al., *Effect of Screening With Primary Cervical HPV Testing vs Cytology Testing on High-grade Cervical Intraepithelial Neoplasia at 48 Months: The HPV FOCAL Randomized Clinical Trial.* JAMA, 2018. **320**(1): p. 43–52.
75. Ronco, G., et al., *Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials.* Lancet, 2014. **383**(9916): p. 524–32.
76. Meijer, C.J., et al., *Guidelines for human papillomavirus DNA test requirements for primary cervical cancer screening in women 30 years and older.* Int J Cancer, 2009. **124**(3): p. 516–20.
77. Arbyn, M., et al., *VALGENT: A protocol for clinical validation of human papillomavirus assays.* J Clin Virol, 2016. **76 Suppl 1**: p. S14-S21.
78. Arbyn, M., et al., *Which high-risk HPV assays fulfil criteria for use in primary cervical cancer screening?* Clin Microbiol Infect, 2015. **21**(9): p. 817–26.
79. Bauer, H.M., Greer, C.E., Manos, M.M., *Determination of genital human papillomavirus infection using consensus PCR*, in *Diagnostic molecular pathology: a practical approach*. 1992. p. 132–152.
80. Gravitt, P.E., et al., *Improved amplification of genital human papillomaviruses.* J Clin Microbiol, 2000. **38**(1): p. 357–61.
81. Xu, L., et al., *Assessment of the Roche Linear Array HPV Genotyping Test within the VALGENT framework.* J Clin Virol, 2018. **98**: p. 37–42.
82. Jacobs, M.V., et al., *A general primer GP5+/GP6(+)-mediated PCR-enzyme immunoassay method for rapid detection of 14 high-risk and 6 low-risk human papillomavirus genotypes in cervical scrapings.* J Clin Microbiol, 1997. **35**(3): p. 791–5.
83. de Roda Husman, A.M., et al., *The use of general primers GP5 and GP6 elongated at their 3' ends with adjacent highly conserved sequences improves human papillomavirus detection by PCR.* J Gen Virol, 1995. **76 (Pt 4)**: p. 1057–62.
84. Heideman, D.A., et al., *Clinical validation of the cobas 4800 HPV test for cervical screening purposes.* J Clin Microbiol, 2011. **49**(11): p. 3983–5.

85. Kleter, B., et al., *Development and clinical evaluation of a highly sensitive PCR-reverse hybridization line probe assay for detection and identification of anogenital human papillomavirus.* J Clin Microbiol, 1999. **37**(8): p. 2508–17.
86. Kleter, B., et al., *Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses.* Am J Pathol, 1998. **153**(6): p. 1731–9.
87. Winer, R.L., et al., *Viral load and short-term natural history of type-specific oncogenic human papillomavirus infections in a high-risk cohort of midadult women.* Int J Cancer, 2014. **134**(8): p. 1889–98.
88. Trevisan, A., et al., *Human papillomavirus type 16 viral load measurement as a predictor of infection clearance.* J Gen Virol, 2013. **94**(Pt 8): p. 1850–7.
89. Snijders, P.J., et al., *Determination of viral load thresholds in cervical scrapings to rule out CIN 3 in HPV 16, 18, 31 and 33-positive women with normal cytology.* Int J Cancer, 2006. **119**(5): p. 1102–7.
90. Apter, D., et al., *Efficacy of human papillomavirus 16 and 18 (HPV-16/18) AS04-adjuvanted vaccine against cervical infection and precancer in young women: final event-driven analysis of the randomized, double-blind PATRICIA trial.* Clin Vaccine Immunol, 2015. **22**(4): p. 361–73.
91. Brown, D.R., et al., *The impact of quadrivalent human papillomavirus (HPV; types 6, 11, 16, and 18) L1 virus-like particle vaccine on infection and disease due to oncogenic nonvaccine HPV types in generally HPV-naive women aged 16–26 years.* J Infect Dis, 2009. **199**(7): p. 926–35.
92. Joura, E.A., et al., *A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women.* N Engl J Med, 2015. **372**(8): p. 711–23.
93. Donken, R., et al., *High effectiveness of the bivalent HPV vaccine up to six years post-vaccination against incident and persistent HPV infections in young Dutch females.* J Infect Dis, 2018. **217**(10): p. 1579–1589.
94. Woestenberg, P.J., et al., *Bivalent Vaccine Effectiveness Against Type-Specific HPV Positivity: Evidence for Cross-Protection Against Oncogenic Types Among Dutch STI Clinic Visitors.* J Infect Dis, 2018. **217**(2): p. 213–222.
95. Lipsitch, M., *Interpreting results from trials of pneumococcal conjugate vaccines: a statistical test for detecting vaccine-induced increases in carriage of nonvaccine serotypes.* Am J Epidemiol, 2001. **154**(1): p. 85–92.
96. Tota, J.E., et al., *Epidemiologic approaches to evaluating the potential for human papillomavirus type replacement postvaccination.* Am J Epidemiol, 2013. **178**(4): p. 625–34.
97. Murall, C.L., K.S. McCann, and C.T. Bauch, *Revising ecological assumptions about Human papillomavirus interactions and type replacement.* J Theor Biol, 2014. **350**: p. 98–109.
98. Gray, P., et al., *Evaluation of HPV type-replacement in unvaccinated and vaccinated adolescent females-Post-hoc analysis of a community-randomized clinical trial (II).* Int J Cancer, 2018. **142**(12): p. 2491–2500.
99. Mori, S., et al., *Biased amplification of human papillomavirus DNA in specimens containing multiple human papillomavirus types by PCR with consensus primers.* Cancer Sci, 2011. **102**(6): p. 1223–7.
100. Kavanagh, K., et al., *Introduction and sustained high coverage of the HPV bivalent vaccine leads to a reduction in prevalence of HPV 16/18 and closely related HPV types.* Br J Cancer, 2014. **110**(11): p. 2804–11.
101. Markowitz, L.E., et al., *Reduction in human papillomavirus (HPV) prevalence among young women following HPV vaccine introduction in the United States, National Health and Nutrition Examination Surveys, 2003–2010.* J Infect Dis, 2013. **208**(3): p. 385–93.
102. Tota, J.E., et al., *Evaluation of human papillomavirus type replacement postvaccination must account for diagnostic artifacts: masking of HPV52 by HPV16 in anogenital specimens.* Cancer Epidemiol Biomarkers Prev, 2015. **24**(1): p. 286–90.

103. zur Hausen, H., *Papillomaviruses in the causation of human cancers–a brief historical account.* Virology, 2009. **384**(2): p. 260–5.
104. Palmer, A.K., A.L. Harris, and R.M. Jacobson, *Human papillomavirus vaccination: a case study in translational science.* Clin Transl Sci, 2014. **7**(5): p. 420–4.
105. Zhou, J., et al., *Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is sufficient for assembly of HPV virion-like particles.* Virology, 1991. **185**(1): p. 251–7.
106. Kirnbauer, R., et al., *Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic.* Proc Natl Acad Sci U S A, 1992. **89**(24): p. 12180–4.
107. Kirnbauer, R., et al., *Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles.* J Virol, 1993. **67**(12): p. 6929–36.
108. Garland, S.M., et al., *Natural history of genital warts: analysis of the placebo arm of 2 randomized phase III trials of a quadrivalent human papillomavirus (types 6, 11, 16, and 18) vaccine.* J Infect Dis, 2009. **199**(6): p. 805–14.
109. Harper, D.M. and L.R. DeMars, *HPV vaccines–A review of the first decade.* Gynecol Oncol, 2017. **146**(1): p. 196–204.
110. Group, F.I.S., *Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions.* N Engl J Med, 2007. **356**(19): p. 1915–27.
111. Brotherton, J.M.L. and P.N. Bloem, *Population-based HPV vaccination programmes are safe and effective: 2017 update and the impetus for achieving better global coverage.* Best Pract Res Clin Obstet Gynaecol, 2018. **47**: p. 42–58.
112. Harper, D.M., *Impact of vaccination with Cervarix (trade mark) on subsequent HPV-16/18 infection and cervical disease in women 15–25 years of age.* Gynecol Oncol, 2008. **110**(3 Suppl 1): p. S11–7.
113. Naud, P.S., et al., *Sustained efficacy, immunogenicity, and safety of the HPV-16/18 AS04-adjuvanted vaccine: final analysis of a long-term follow-up study up to 9.4 years post-vaccination.* Hum Vaccin Immunother, 2014. **10**(8): p. 2147–62.
114. Paavonen, J., et al., *Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women.* Lancet, 2009. **374**(9686): p. 301–14.
115. Skinner, S.R., et al., *Human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine for the prevention of cervical cancer and HPV-related diseases.* Expert Rev Vaccines, 2016. **15**(3): p. 367–87.
116. Struyf, F., et al., *Post hoc analysis of the PATRICIA randomized trial of the efficacy of human papillomavirus type 16 (HPV-16)/HPV-18 AS04-adjuvanted vaccine against incident and persistent infection with nonvaccine oncogenic HPV types using an alternative multiplex type-specific PCR assay for HPV DNA.* Clin Vaccine Immunol, 2015. **22**(2): p. 235–44.
117. Wheeler, C.M., et al., *The impact of quadrivalent human papillomavirus (HPV; types 6, 11, 16, and 18) L1 virus-like particle vaccine on infection and disease due to oncogenic nonvaccine HPV types in sexually active women aged 16–26 years.* J Infect Dis, 2009. **199**(7): p. 936–44.
118. Schiller, J. and D. Lowy, *Explanations for the high potency of HPV prophylactic vaccines.* Vaccine, 2018.
119. Herrero, R., et al., *Prevention of persistent human papillomavirus infection by an HPV16/18 vaccine: a community-based randomized clinical trial in Guanacaste, Costa Rica.* Cancer Discov, 2011. **1**(5): p. 408–19.

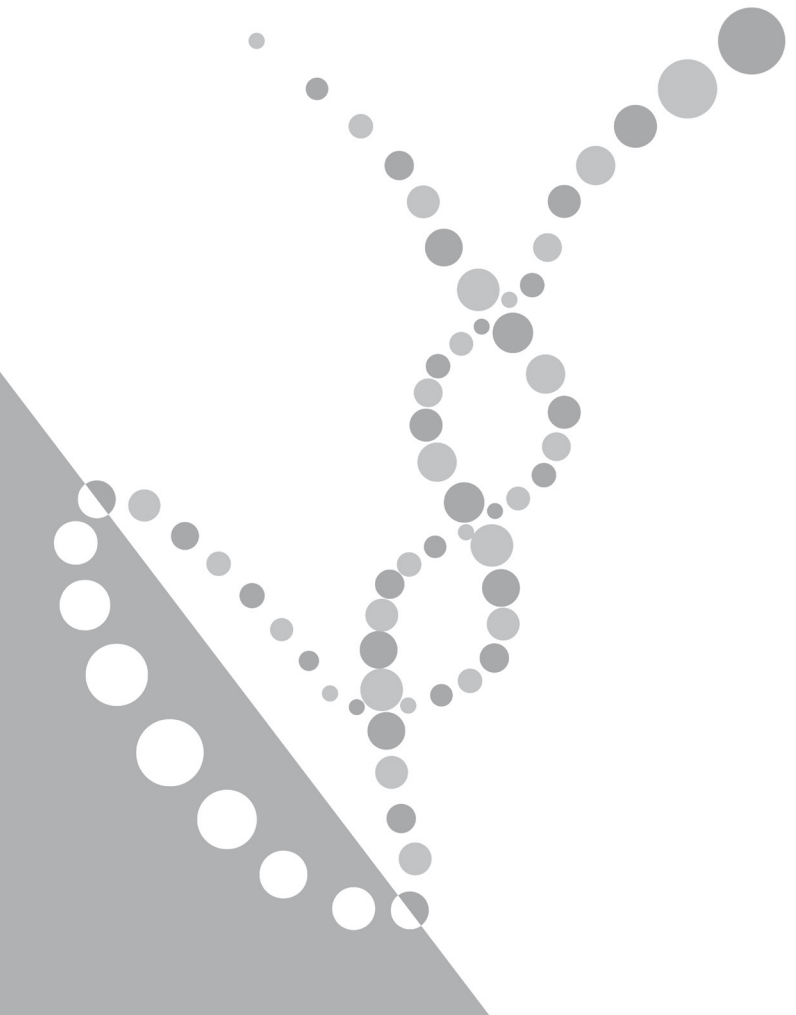
120. Einstein, M.H., et al., *Comparative humoral and cellular immunogenicity and safety of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine and HPV-6/11/16/18 vaccine in healthy women aged 18–45 years: follow-up through Month 48 in a Phase III randomized study.* Hum Vaccin Immunother, 2014. **10**(12): p. 3455–65.
121. Schurink-Van't Klooster, T.M., et al., *HPV vaccination: Background information for the Dutch Health Council.* RIVM, 2017.
122. IARC HPV Working Group, *Primary End-points for Prophylactic HPV Vaccine Trials*. Primary End-points for Prophylactic HPV Vaccine Trials. 2014, Lyon (FR).
123. Harper, D.M., et al., *Sustained efficacy up to 4.5 years of a bivalent L1 virus-like particle vaccine against human papillomavirus types 16 and 18: follow-up from a randomised control trial.* Lancet, 2006. **367**(9518): p. 1247–55.
124. Malagon, T., et al., *Cross-protective efficacy of two human papillomavirus vaccines: a systematic review and meta-analysis.* Lancet Infect Dis, 2012. **12**(10): p. 781–9.
125. Szarewski, A., et al., *Efficacy of the HPV-16/18 AS04-adjuvanted vaccine against low-risk HPV types (PATRICIA randomized trial): an unexpected observation.* J Infect Dis, 2013. **208**(9): p. 1391–6.
126. Howell-Jones, R., et al., *Declining genital Warts in young women in england associated with HPV 16/18 vaccination: an ecological study.* J Infect Dis, 2013. **208**(9): p. 1397–403.
127. Woestenberg, P.J., et al., *No evidence for cross-protection of the HPV-16/18 vaccine against HPV-6/11 positivity in female STI clinic visitors.* J Infect, 2017. **74**(4): p. 393–400.
128. Giannini, S.L., et al., *Enhanced humoral and memory B cellular immunity using HPV16/18 L1 VLP vaccine formulated with the MPL/aluminium salt combination (AS04) compared to aluminium salt only.* Vaccine, 2006. **24**(33–34): p. 5937–49.
129. Harari, A., et al., *Cross-protection of the Bivalent Human Papillomavirus (HPV) Vaccine Against Variants of Genetically Related High-Risk HPV Infections.* J Infect Dis, 2016. **213**(6): p. 939–47.
130. Mollers, M., et al., *Effectiveness of human papillomavirus vaccine against incident and persistent infections among young girls: Results from a longitudinal Dutch cohort study.* Vaccine, 2015. **33**(23): p. 2678–83.
131. Mollers, M., et al., *Prevalence of genital HPV infections and HPV serology in adolescent girls, prior to vaccination.* Cancer Epidemiol, 2012. **36**(6): p. 519–24.
132. Romanowski, B., et al., *Immunogenicity and safety of the HPV-16/18 AS04-adjuvanted vaccine administered as a 2-dose schedule compared with the licensed 3-dose schedule: results from a randomized study.* Hum Vaccin, 2011. **7**(12): p. 1374–86.
133. Romanowski, B., et al., *Immune response to the HPV-16/18 AS04-adjuvanted vaccine administered as a 2-dose or 3-dose schedule up to 4 years after vaccination: results from a randomized study.* Hum Vaccin Immunother, 2014. **10**(5): p. 1155–65.
134. Romanowski, B., et al., *Sustained immunogenicity of the HPV-16/18 AS04-adjuvanted vaccine administered as a two-dose schedule in adolescent girls: Five-year clinical data and modeling predictions from a randomized study.* Hum Vaccin Immunother, 2016. **12**(1): p. 20–9.
135. Donken, R., et al., *Immune Responses After 2 Versus 3 Doses of HPV Vaccination up to 4(1/2) Years After Vaccination: An Observational Study Among Dutch Routinely Vaccinated Girls.* J Infect Dis, 2017. **215**(3): p. 359–367.
136. de Kok, I.M., et al., *Trends in cervical cancer in the Netherlands until 2007: has the bottom been reached?* Int J Cancer, 2011. **128**(9): p. 2174–81.

137. Nanda, K., et al., *Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review.* Ann Intern Med, 2000. **132**(10): p. 810–9.
138. Kitchener, H.C., P.E. Castle, and J.T. Cox, *Chapter 7: Achievements and limitations of cervical cytology screening.* Vaccine, 2006. **24 Suppl 3**: p. S3/63–70.
139. Polman, N.J., et al., *HPV-based cervical screening: Rationale, expectations and future perspectives of the new Dutch screening programme.* Prev Med, 2019. **119**: p. 108–117.
140. Elfstrom, K.M., et al., *Long term duration of protective effect for HPV negative women: follow-up of primary HPV screening randomised controlled trial.* BMJ, 2014. **348**: p. g130.
141. RIVM. *Bevolkingsonderzoek baarmoederhalskanker*. Available from: https://www.rivm.nl/bevolking-sonderzoek-baarmoederhalskanker.
142. Melnikow, J., et al., *Cervical intraepithelial neoplasia outcomes after treatment: long-term follow-up from the British Columbia Cohort Study.* J Natl Cancer Inst, 2009. **101**(10): p. 721–8.
143. Kocken, M., et al., *Risk of recurrent high-grade cervical intraepithelial neoplasia after successful treatment: a long-term multi-cohort study.* Lancet Oncol, 2011. **12**(5): p. 441–50.
144. Kocken, M., et al., *High-risk human papillomavirus testing versus cytology in predicting post-treatment disease in women treated for high-grade cervical disease: a systematic review and meta-analysis.* Gynecol Oncol, 2012. **125**(2): p. 500–7.
145. Integraal Kankercentrum Nederland. *Richtlijn CIN, AIS EN VAIN*. 2015; Available from: https://www.on-coline.nl/cin-ais-en-vain.

Part I

Viral load measurements in vaccinated and non-vaccinated settings

Chapter 2



Correlation between viral load, multiplicity of infection, and persistence of HPV16 and HVP18 infection in a dutch cohort of young women

Pascal van der Weele[a,b,], Elske van Logchem[a], Petra Wolffs[c,1], Ingrid van den Broek[a,2] Mariet Feltkamp[d], Hester de Melker[a], Chris JLM Meijer[b], Hein Boot[a,3] and Audrey King[a]

[a]National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven, the Netherlands
[b]Vrije Universiteit - University Medical Center (VUmc), Department of Pathology, Amsterdam, the Netherlands
[c]Maastricht University Medical Center (MUMC+), Department of Medical Microbiology, Maastricht, The Netherlands
[d]Leiden University Medical Center (LUMC), Department of Medical Microbiology, Leiden, the Netherlands
[1]on behalf of the Medical Microbiological Laboratories
[2]on behalf of the CSI group
[3]Deceased

J Clin Virol. 2016 Oct;83:6-11

Abstract

Background Persistent high-risk human papillomavirus (hrHPV) infection precedes the development of cervical cancer. Here we evaluated the contribution of HPV16/18 viral load and the presence of infections with multiple HPV types to persistence and clearance of HPV16/18 infections.
Methods Vaginal self-swabs were obtained from young women (16–29y) with one year interval. HPV genotyping was performed using the highly sensitive SPF10-DEIA-LiPA$_{25}$ system. HPV16/18 DNA loads were quantified via an adapted, highly sensitive qPCR protocol targeting the L1 gene.
Results We identified 227 HPV16 and 111 HPV18 infections with follow-up. For HPV16 132/227 (58%) were persistent and 95/227 cleared. For HPV18 49/111 (44%) infections were persistent and 62/111 cleared. Baseline viral load was significantly higher in persistent infections than in clearing infections for both HPV16 ($p = 0.022$) and HPV18 ($p = 0.013$). At baseline, only HPV16 viral load was significantly higher in multiple HPV infections compared with single infections ($p = 0.003$). In logistic regression analysis HPV16 and HPV18 viral load were found to contribute to persistency with OR = 1.279 (95%CI = 1.074–1.524) and OR = 1.256 (95%CI = 1.028–1.533) per log-unit increase HPV16 and HPV18 viral load respectively. The presence of multiple HPV type infections was not associated with higher persistency.
Conclusion HPV16/18 viral load might be used as a marker for persisting infections and is affected by the presence of multiple HPV infections. Evaluation of these parameters at the population level may be of value to assess the presence of persistent or clearing HPV16/18 infections as an early marker, and may provide useful quantitative information in (epidemiological) vaccine monitoring studies.
Keywords HPV16, HPV18, Persistent infection, Viral load, Multiple HPV infection

Introduction

A persistent high risk HPV type (hrHPV) infection is required for the development of precancerous lesions and ultimately cervical cancer [1]. HPV16 and HPV18 are the most prevalent hrHPVs causing cervical cancer [2]. Infections with hrHPV are common and generally transient [3, 4]. About 20% of hrHPV infections persist and can develop into cervical intraepithelial neoplasia 1–3 (CIN1–3) ultimately leading up to cervical cancer [5]. The duration of this development can take 15–30 years in total [6, 7].

High viral load (VL) has been described as a contributing factor for infection persistence and the development of CIN and cancer. Higher HPV16 and HPV18 VLs have been found in women with CIN lesions compared to women with normal cytology or without CIN lesions [8, 9]. Although discriminating in lesion severity proved elusive [10–13], a VL threshold value can be set to assess risk of CIN3 [9]. For persisting HPV infections, VL has been modeled to predict progressive CIN disease [14]. Indeed, in a recent study targeting HPV infection in mid-adult women it has been shown that increased HPV VL leads to significantly higher odds of persistent infections [15]. Conversely, low HPV16 and HPV18 VL have been associated with clearance of HPV infection [16, 17]. Additionally, multiple HPV infections have been associated with increased risk of CIN2–3 [18], and increased odds of short-term persistency [19]. However, individual HPV types in multiple HPV infections act independent of each other in lesions [20] and the increased risk of CIN2+ in multiple infections should be interpreted as a cumulative effect rather than a synergistic effect [21]. As a result, currently no clinical indication for the use of VL or multiple HPV infections in a clinical setting have been established.

Prophylactic HPV16/18 vaccines have been introduced to prevent vaccine-type related cervical cancer. Limited cross-protective effects against phylogenetically related non-vaccine HPV types has been described [22]. To assess HPV vaccine effectiveness in pre-adolescent women intermediate or surrogate markers are used. Persistent HPV infection is one of the first outcomes to measure the potential effect of vaccination. HPV vaccines have shown to be effective against persisting infections and vaccine type associated CIN2+ lesions [23–25]. The International Agency for Research on Cancer (IARC) and the World Health Organization (WHO) have stated in 2014 that persistent HPV infection is now considered a valid HPV vaccination end-point. Thus discriminating between persistent and clearing HPV infections may be valuable in assessing vaccine effectiveness**.**

Objectives

We assessed whether VL measurements at baseline can differentiate between persisting and clearing HPV16/18 infections in young women (aged 16–29y) using a highly sensitive VL assay. Additionally the effect of multiple HPV-type infections on persistence of HPV16/18 infections is analyzed. Finally, contributions of HPV16 and HPV18 VL and multiple HPV-type infections towards infection persistency were studied.

Materials and Methods

Study design and HPV DNA detection and genotyping

Vaginal self-swabs were obtained from the *Chlamydia trachomatis* Screening Implementation program (CSI-study) and stored at -20°C until processing. Recruitment criteria, study methods and statistics have been described before [26–29]. Out of 3282 initial participants, 2014 participated in round two (61.4%). The mean time between sampling rounds was 50.4 weeks (5 – 101 weeks). DNA was isolated from swabs and tested for HPV DNA using SPF10 PCR [30, 31] (DDL Diagnostics Laboratory, Voorburg, The Netherlands) as described previously [28]. HPV presence was assessed via HPV DNA Enzyme ImmunoAssay (DEIA). DEIA-positive samples were genotyped through a reverse hybridization line probe assay (LiPA) [31] and scored by visual inspection, allowing detection of 25 HPV genotypes [32]. Persistent infections were defined as type-specific HPV genotyping results in two subsequent samples. Clearing infections were defined as loss of type-specific HPV detection in a follow-up sample. Samples without available follow-up, or with follow-up <26 weeks, were omitted from analyses comparing persisting and clearing infections. Single HPV type infections were defined as samples that were positive for a single HPV type. Multiple HPV-type infections were defined as detection of one or more HPV genotypes besides HPV16 or HPV18.

Quantification of relative HPV-DNA loads by Real-Time PCR

VL was determined for 383 first round HPV16 (n = 271) and HPV18-positive (n = 112) participants (Figure 1) using an L1-targeting quantitative (q)PCR protocol adapted from Seaman *et al.* [33] which was optimized to approach SPF10-LiPA$_{25}$ sensitivity levels. qPCRs were performed in 20 µl final volume using LightCycler TaqMan Master on the Roche LightCycler 480 platform (Roche Diagnostics, Almere, the Netherlands). An in-sequence 5' thymine residue was added to the HPV16L1Probe. The HPV18L1Probe 5' HEX label was replaced with a 5' YakimaYellow label. qPCRs were performed in singleplex to enhance sensitivity. Primers were directed against 136 bp HPV16 and 120 bp HPV18 L1 fragments (Table 1). Cycling conditions consisted of a 95°C denaturation step for 10 minutes, followed by 50 cycles consisting of alternating 95°C for 15 seconds, and 60°C for 30 seconds. Fluorescence was detected after every elongation step.

In each run a standard plasmid, containing HPV16 or HPV18 DNA was included. Plasmids were diluted to 3.13×10^3 Genome Equivalents (GEq) and spiked with human genomic DNA (Roche Diagnostics, Almere, The Netherlands) at 200 ng/reaction to determine absolute sample concentrations. VLs were determined based on cycle threshold (Ct) values above their corresponding baselines in relation to the quantitative standards.

To normalize HPV copies to the amount of human genomic DNA, beta-actin quantifications were included. Beta-actin specific primers (Table 1) were used in a qPCR with conditions according to [34]. Quantities of human DNA determined by the assay were converted to number of cells

assuming 150.000 human cells per µg of human genomic DNA [35]. DNA loads were expressed as the number of HPV DNA copies relative to the amount of human material, genome equivalents/million human cells (GEq/mhc). Samples that were SPF10-LiPA$_{25}$ positive, but negative in VL assay were assigned an arbitrary concentration of 1 GEq/mhc.

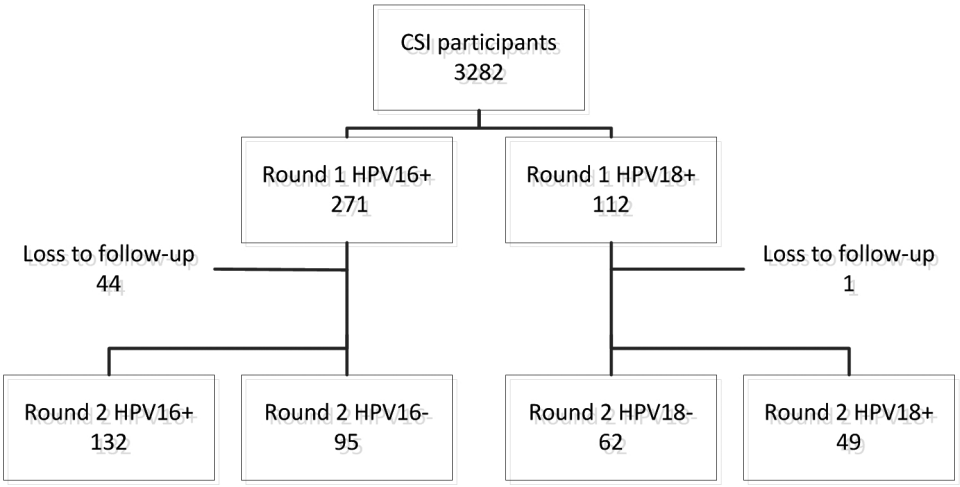


Figure 1: Schematic representation the CSI study for rounds one andtwo. In round one only HPV16+ and HPV18+ participants with VL information were included.

Standard curves

To facilitate relative quantification of HPV16 and HPV18 DNA in samples, serial dilutions of plasmids containing full-length HPV16 or HPV18 were tested. Standardized plasmids were obtained from the 1st WHO International Standard for Human Papillomavirus (NIBSC: 06/202, 06/206). Twofold serial dilutions were made with concentrations ranging from $5.0x10^{4}$–$1.5x10^{0}$ GEq/reaction. A beta-actin standard curve with concentrations ranging from $1.0x10^{3}$-$2.0x10^{-5}$ng/reaction was obtained by making fivefold dilutions of human genomic DNA (Roche), to quantify cellular material in each sample. Detection limits were determined by plotting Ct values against sample concentration [33] (Figure S1). Repeat testing of HPV16 and HPV18 plasmid DNA in twofold serial dilutions showed detection limits of 5 GEq/reaction (1 GEq/µl) for both HPV16 and HPV18. Detection up to 1 GEq/µl occurs in a linear fashion for both HPV16 ($R^2 = 0.99$) and HPV18 ($R^2 = 0.99$, Figure S1). No cross-reactivity was found for HPV16 or HPV18 with plasmids containing HPV18/16, 31, 33, 45, 51, 52 and 58 at concentrations around 10^{5} GEq/µl (Figure S2). The addition of plasmids showed no effect in quantitative results for HPV16 or HPV18 ($R^2 = 1.0$, $R^2 = 0.997$, data not shown).

Table 1: Sequences and concentration of HPV16, HPV18 [33] and beta-actin primers [34] used in qPCR. In bold are adaptations to the protocol as described by [33].

Oligonucleotide	Sequence	Concentration
HPV16L1F	5'-TTGTTGGGGTAACCAACTATTTGTTACTGTT-3'	400 nM
HPV16L1R	5'-CCTCCCCATGTCTGAGGTACTCCTTAAAG-3'	400 nM
HPV16L1Probe	6FAM-5'-**T**GTCATTATGTGCTGCCATATCTACTTC-3'-TAMRA	**200 nM**
HPV18L1F	5'-GCATAATCAATTATTTGTTACTGTGGTAGATACCACT-3'	400 nM
HPV18L1R	5'-GCTATACTGCTTAAATTTGGTAGCATCATATTGC-3'	400 nM
HPV18L1Probe	**YakYellow**-5'-AACAATATGTGCTTCTACACAGTCTCCTGT-3'-BHQ2	**200 nM**
Beta-Actin F	5'-TCACCCACACTGTGCCCATCTACGA-3'	1000 nM
Beta-Actin R	5'-CAGCGGAACCGCTCATTGCCAATGG-3'	1000 nM
Beta-Actin Probe	6FAM-5'-ATGCCCTCCCCCATGCCATCCTGCGT-3'-TAMRA	500 nM

Statistical analyses

SPSS Statistics 19 was used to perform statistical analysis. To assess a difference in VL between persisting and clearing infections the Mann-Whitney U (MWU) test was employed. Samples positive for HPV16 or HPV18 in the first round without follow-up data were omitted. Two-sided p-values of < 0.05 were considered statistically significant. Receiver operator characteristic (ROC) curves with area under the curve (AUC) were calculated to determine the discriminatory value of VL for persistent infections. Binary logistical regression was used to analyze if specific conditions were associated with risk of persistent infections and interaction terms were included to distinguish individual from combined parameter effects. Frequency tables were analyzed using Fisher's exact test. Odds ratios (OR) were calculated with 95% confidence intervals (CI). CIs not overlapping OR = 1 were considered to be statistically significant. GraphPad 7.0 was used to generate a scatter dot plot of the VL data.

Results

Study characteristics

Of 271 HPV16 infections detected in the first round, 44 were lost to follow-up (Figure 1). Of the remaining 227 HPV16 infections, 132 persisted (58.1%) in round two and 95 infections cleared. Of 112 HPV18 infections detected in the first round, one was excluded and 49 persisted (44.1%) in round two, whereas 62 infections were cleared.

HPV16/18 VLs in persistent and clearing infections

We introduce enhanced HPV16 and HPV18 qPCRs detection limits in line with SPF10 PCR based HPV DNA detection levels. Scatter dot plots of VL spread for HPV16 and HPV18 in persistent and clearing infections are displayed in Figure 2. HPV16 VL was found to be significantly higher in initial samples of the 132 participants with persistent infections compared to 95 participants with clearing infections (MWU, p = 0.022) (Table 2). A similar significant correlation was observed for HPV18, with higher VLs detected more frequently among participants with persistent HPV18 infections compared to clearing infections (MWU, p = 0.013) (Table 2). Similar significancies were obtained when performing the same analyses on raw data not corrected for total cell content (p = 0.013 for HPV16 and p = 0.035 for HPV18). ROC analysis shows areas under the curve (AUC) of 0.59 for HPV16 and 0.64 for HPV18 (data not shown).

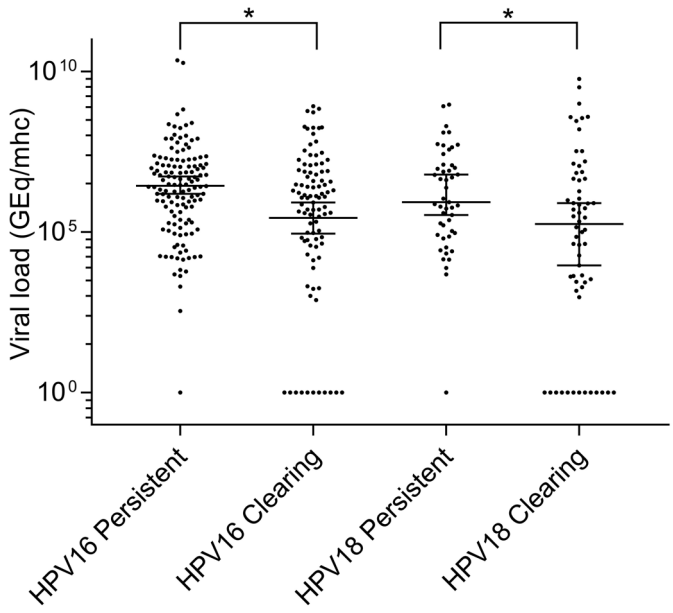


Figure 2: Scatter dot plots of VL measurements for HPV16 and HPV18, classified as persisting or clearing infections. SPF10 positive VL negative samples were assigned an arbitrary concentration of 1 GEq/mhc.

Table 2: HPV16/18 DNA VL in persistent, clearing, single HPV and multiple HPV infections. VLs are expressed in genome equivalents per million human cells (GEq/mhc. Loss to followup was irrelevant for the multiplicity analysis.

	HPV16	HPV18
Total (n)	271	111
Loss to followup, n	44	1
Persistent, n (%)	132 (58.1%)	49 (44.1%)
Median load (95% CI)	$2.80x10^6$ ($1.56x10^6$ – $5.48x10^6$)	$8.57x10^5$ ($3.41x10^5$ – $6.29x10^6$)
Clearing, n	95	62
Median load (95% CI)	$1.19x10^6$ ($4.04x10^5$ – $2.51x10^6$)	$1.80x10^5$ ($9.16x10^3$ – $8.07x10^5$)
Significance (p)	0.022	0.013
Single infections, n (%)	81 (29.9%)	19 (17.0%)
Median load (95% CI)	$7.08x10^5$ ($2.10x10^5$ – $1.54x10^6$)	$1.08x10^5$ ($1.00x10^0$ – $7.44x10^6$)
Multiple HPV, n (%)	190 (70.1%)	93 (83.0%)
Median load (95% CI)	$2.77x10^6$ ($1.57x10^6$ – 5.16x106)	$6.86x10^5$ ($1.88x10^5$ – $1.89x10^6$)
Significance (p)	0.003	0.129

Association of multiple HPV-type infection and HPV16/18 VL

In 271 women tested positive for HPV16 in the initial round, 81 (29.9%) were positive for HPV16 only. In 190 women (70.1%), other HPV types were detected in addition to HPV16.

HPV16 VL was found to be significantly higher in the participants with multiple HPV status compared to the women with HPV16 single infections (MWU, $p = 0.003$) (Table 2).

Out of 112 women with baseline HPV18 infections, 19 (17.0%) had HPV18 single infections and 93 (83.0%) had multiple HPV infections. For HPV18, a similar, but non-significant trend was observed between VL and multiple HPV status (MWU, $p = 0.129$) (Table 2).

Both HPV16 and HPV18 multiple infections are not more likely to persist than single infections (Fisher's exact test, $p = 0.240$ and $p = 0.311$).

Modeling contribution of VL and multiple HPV status to persistency

Correlations with HPV16/18 persistency were found for HPV16/18 VL and multiple HPV-type infections. Integrated in a binary logistic regression model their combined effect on the onset of persistency can be assessed. Persistency is defined as the dependent, dichotomous outcome variable. VL and multiple HPV status were defined as independent continuous and dichotomous variables respectively.

For both HPV16 and HPV18, VL contributed significantly to the outcome of persistency (p = 0.006 and p = 0.026). HPV16 infections were 27.9% more likely to persist per log-unit increase of VL (OR = 1.279, CI = 1.074–1.524)and HPV18 infections were 25.6% more likely to persist per log-unit increase of VL (OR = 1.256, CI = 1.028–1.533). The interaction effect between VL and multiplicity of infection was not significant for HPV16 (p = 0.944) or HPV18 (p = 0.770). In this model, the effect of multiplicity of infection on infection persistence was not significant for HPV16 (0.779) or HPV18 (p = 0.600), implying no effect of multiple HPV infection on HPV16/18 infection persistence for this dataset.

Discussion

Using a highly sensitive VL assay we show significantly higher HPV16 and HPV18 loads in women with persistent infections compared to women who cleared their infections (MWU, p = 0.022 and p = 0.013). These findings are in agreement with literature [8, 9, 15], however, this is the first time this correlation was also demonstrated in young women combined with sensitive quantification of viral load. We have applied an adapted quantitative assay in line with sensitivity demonstrated by SPF10 PCR based genotyping assay, which is potentially able to detect HPV infections at a very early stage. The optimized, singleplex assays show detection limits of 5 GEq/reaction (1 GEq/µl) for both HPV16 and HPV18; fourfold lower than the original assay [33]. These adaptations make the assays more suitable for the quantification of HPV DNA load from infections detected with SPF10-LiPA$_{25}$ platform used in our studies.This platform is considered to be a factor ten more sensitive on average than the GP5+/6+ and (PG)MY PCR based HPV tests [36]. The increased sensitivity of the assays was of particular use within the scope of this research, as it is focused on a population of young women potentially targeted for HPV vaccination. It should be noted that overdetection of HPV infections with low VL may limit the use of our VL assays, because most of these infections are not expected to persist. However, Figure 2 shows that some low VL infections can persist as well for both HPV16 and HPV18, highlighting the usefullness of high sensitivity testing.Our findings suggest a biological correlation between VL and persistent infections in women with HPV16/18 infections at an earlier stage of infection than described previously. ROC analysis shows that although the difference in median VL between persistent and clearing HPV infections is significant at the population level; HPV16/18 VL on its own is insufficient to predict persistent infections at the individual level (AUC values of 0.595 for HPV16 and 0.630 for HPV18). Whereas the median HPV16/18 VL in persisting infections is higher than in clearing infections, the variation in VL overlaps, as illustrated by Figure 2. This makes the setting of a cut-off to exclude clearing infection with low VL impossible.

In addition to the correlation found between HPV16/18 VL and persistent infections, significantly higher HPV16 VL levels were found in multiple HPV-type infections compared to single HPV16 infections (MWU, p = 0.003) (Table 2). Significance was not reached for HPV18 (MWU, p = 0.129), probably due to the small number of single HPV18 infections in our study (n = 19).

Interestingly, the higher HPV16/18 VL levels in multiple HPV-type infections were not associated with significantly more persistent HPV16 or HPV18 infections (Fisher's exact test $p = 0.240$ and $p = 0.311$). This suggests that above certain VL levels persistent infections will occur.

The binary logistic regression analysis gives insight in the contributions of VL and multiple HPV-type infection status towards persistent infections. Both HPV16 and HPV18 VL levels were significant contributors to persistent infections. Multiple HPV-type infections possibly convey extra risk of persisting HPV16/18 infections, but within this dataset this was not significant. For HPV18 the small number of single infections could have skewed the analysis. Additional research could explain the behavior of VL in multiple HPV-type infections and the contribution to onset of persisting infections.

Although baseline VL measurements cannot be used to predict persisting infections at the individual level, the correlation between VL and persisting infections might be useful to predict an increase or decrease in persistent HPV infections at the population level. Thus, evaluation of HPV16/18 VL in cervical samples from (epidemiological) studies comparing vaccinated to unvaccinated cohorts of women might be used to show a decrease in the median VL in the vaccinated cohort. This could then be interpreted as a proxy for a decrease of persistent infections in the vaccinated cohort.

This study has some limitations as it was originally set up for *Chlamydia trachomatis* (*Ct*) screening and adapted for HPV purposes. Originally the study spanned four rounds, but over the first three rounds a 41% decline in participation for the *Ct* screening was found [27]. Upon adaptation for HPV the loss to follow-up increased to 72% over three rounds. The initial rounds provided sufficient samples for analysis, but numbers from the following rounds were too small to reach statistical significance. As a result, rounds three and four were omitted from this analysis. In contrast to the HPV16 data, we found no loss to follow-up among women with HPV18. One sample was excluded as the time between rounds was < 6 months. Since we obtained similar data for both HPV16 and HPV18, we do not believe this discrepancy influenced our results. Ideally, persisting infections would be defined as two positive type-specific measurements preceded by a negative baseline. Due to loss to follow-up, we were unable to utilize first round information as such. Because of this concession, the group of clearing infections consisted of a mix of incident, transient, and clearing persistent infections as no information is known before the first measurement. Despite this possible drawback our findings are in line with previous results described by Winer *et al.* [15] for midadult women.

Conclusion

We describe correlations between HPV16 and HPV18 VL levels, multiple HPV-type infections and persisting HPV infections in a cohort of young women (aged 16–29y) utilizing highly sensitive HPV16 and HPV18 VL assays, which allow quantification of HPV infections at an early stage. VL levels cannot be used at the individual level to predict persistent HPV infections, however, our findings suggest that median VL levels could represent an early epidemiological marker of persistent HPV16/18 infection in a cohort of young females at population level and, as such, might be of use in (epidemiological) vaccine monitoring studies. In addition, significantly higher VL levels were found for HPV16 in multiple HPV-type infections compared to single HPV16 infections. A similar trend was observed for HPV18. Despite limited predictive value, modeling of VL showed significantly increased odds ratios for persistence per log-unit increase in viral load at the population level.

References

1. Walboomers, J.M., et al., *Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.* J Pathol, 1999. **189**(1): p. 12–9.
2. Arbyn, M., et al., *Are 20 human papillomavirus types causing cervical cancer?* J Pathol, 2014. **234**(4): p. 431–5.
3. Baseman, J.G. and L.A. Koutsky, *The epidemiology of human papillomavirus infections.* J Clin Virol, 2005. **32 Suppl 1**: p. S16–24.
4. Meijer, C.J., P.J. Snijders, and A.J. van den Brule, *Screening for cervical cancer: should we test for infection with high-risk HPV?* CMAJ, 2000. **163**(5): p. 535–8.
5. Ho, G.Y., et al., *Natural history of cervicovaginal papillomavirus infection in young women.* N Engl J Med, 1998. **338**(7): p. 423–8.
6. McCredie, M.R., et al., *Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study.* Lancet Oncol, 2008. **9**(5): p. 425–34.
7. Vink, M.A., et al., *Clinical progression of high-grade cervical intraepithelial neoplasia: estimating the time to preclinical cervical cancer from doubly censored national registry data.* Am J Epidemiol, 2013. **178**(7): p. 1161–9.
8. Hesselink, A.T., et al., *High-risk human papillomavirus DNA load in a population-based cervical screening cohort in relation to the detection of high-grade cervical intraepithelial neoplasia and cervical cancer.* Int J Cancer, 2009. **124**(2): p. 381–6.
9. Snijders, P.J., et al., *Determination of viral load thresholds in cervical scrapings to rule out CIN 3 in HPV 16, 18, 31 and 33-positive women with normal cytology.* Int J Cancer, 2006. **119**(5): p. 1102–7.
10. Briolat, J., et al., *HPV prevalence, viral load and physical state of HPV-16 in cervical smears of patients with different grades of CIN.* Int J Cancer, 2007. **121**(10): p. 2198–204.
11. Xi, L.F., et al., *Effect of cervical cytologic status on the association between human papillomavirus type 16 DNA load and the risk of cervical intraepithelial neoplasia grade 3.* J Infect Dis, 2008. **198**(3): p. 324–31.
12. Lee, S.J., et al., *Semi-quantitative HPV viral load in patients with ASC-US cytology: viral load correlates strongly with the presence of CIN but only weakly with its severity.* Cytopathology, 2014.
13. Schmitt, M., et al., *Viral load of high-risk human papillomaviruses as reliable clinical predictor for the presence of cervical lesions.* Cancer Epidemiol Biomarkers Prev, 2013. **22**(3): p. 406–14.
14. Depuydt, C.E., et al., *Changes in type-specific human papillomavirus load predict progression to cervical cancer.* J Cell Mol Med, 2012. **16**(12): p. 3096–104.
15. Winer, R.L., et al., *Viral load and short-term natural history of type-specific oncogenic human papillomavirus infections in a high-risk cohort of midadult women.* Int J Cancer, 2014. **134**(8): p. 1889–98.
16. Carcopino, X., et al., *Two years outcome of women infected with high risk HPV having normal colposcopy following low-grade or equivocal cytological abnormalities: are HPV16 and 18 viral load clinically useful predictive markers?* J Med Virol, 2012. **84**(6): p. 964–72.
17. Trevisan, A., et al., *Human papillomavirus type 16 viral load measurement as a predictor of infection clearance.* J Gen Virol, 2013. **94**(Pt 8): p. 1850–7.
18. Schmitt, M., et al., *Multiple human papillomavirus infections with high viral loads are associated with cervical lesions but do not differentiate grades of cervical abnormalities.* J Clin Microbiol, 2013. **51**(5): p. 1458–64.

19. Xi, L.F., et al., *Human Papillomavirus (HPV) type 16 and type 18 DNA Loads at Baseline and Persistence of Type-Specific Infection during a 2-year follow-up.* J Infect Dis, 2009. **200**(11): p. 1789–97.
20. Quint, W., et al., *One virus, one lesion—individual components of CIN lesions contain a specific HPV type.* J Pathol, 2012. **227**(1): p. 62–71.
21. Chaturvedi, A.K., et al., *Human papillomavirus infection with multiple types: pattern of coinfection and risk of cervical disease.* J Infect Dis, 2011. **203**(7): p. 910–20.
22. Malagon, T., et al., *Cross-protective efficacy of two human papillomavirus vaccines: a systematic review and meta-analysis.* Lancet Infect Dis, 2012. **12**(10): p. 781–9.
23. Harper, D.M., *Impact of vaccination with Cervarix (trade mark) on subsequent HPV-16/18 infection and cervical disease in women 15–25 years of age.* Gynecol Oncol, 2008. **110**(3 Suppl 1): p. S11–7.
24. Harper, D.M., S.L. Vierthaler, and J.A. Santee, *Review of Gardasil.* J Vaccines Vaccin, 2010. **1**(107).
25. Naud, P.S., et al., *Sustained efficacy, immunogenicity, and safety of the HPV-16/18 AS04-adjuvanted vaccine: final analysis of a long-term follow-up study up to 9.4 years post-vaccination.* Hum Vaccin Immunother, 2014. **10**(8): p. 2147–62.
26. van den Broek, I.V., et al., *Evaluation design of a systematic, selective, internet-based, Chlamydia screening implementation in the Netherlands, 2008–2010: implications of first results for the analysis.* BMC Infect Dis, 2010. **10**: p. 89.
27. van den Broek, I.V., et al., *Effectiveness of yearly, register based screening for chlamydia in the Netherlands: controlled trial with randomised stepped wedge implementation.* BMJ, 2012. **345**: p. e4316.
28. Mollers, M., et al., *Prevalence, incidence and persistence of genital HPV infections in a large cohort of sexually active young women in the Netherlands.* Vaccine, 2013. **31**(2): p. 394–401.
29. van den Broek, I.V., et al., *Systematic selection of screening participants by risk score in a Chlamydia screening programme is feasible and effective.* Sex Transm Infect, 2012. **88**(3): p. 205–11.
30. Kleter, B., et al., *Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses.* Am J Pathol, 1998. **153**(6): p. 1731–9.
31. Kleter, B., et al., *Development and clinical evaluation of a highly sensitive PCR-reverse hybridization line probe assay for detection and identification of anogenital human papillomavirus.* J Clin Microbiol, 1999. **37**(8): p. 2508–17.
32. Humans, I.W.G.o.t.E.o.C.R.t., *Biological agents. Volume 100 B. A review of human carcinogens.* IARC Monogr Eval Carcinog Risks Hum, 2012. **100**(Pt B): p. 1–441.
33. Seaman, W.T., et al., *Detection and quantitation of HPV in genital and oral tissues and fluids by real time PCR.* Virol J, 2010. **7**: p. 194.
34. Zhao, M., et al., *Feasibility of quantitative PCR-based saliva rinse screening of HPV for head and neck cancer.* Int J Cancer, 2005. **117**(4): p. 605–10.
35. Sanchez, J.L. and G.A. Storch, *Multiplex, quantitative, real-time PCR assay for cytomegalovirus and human DNA.* J Clin Microbiol, 2002. **40**(7): p. 2381–6.
36. Snijders, P.J., A.J. van den Brule, and C.J. Meijer, *The clinical relevance of human papillomavirus testing: relationship between analytical and clinical sensitivity.* J Pathol, 2003. **201**(1): p. 1–6.

Supplementary Material

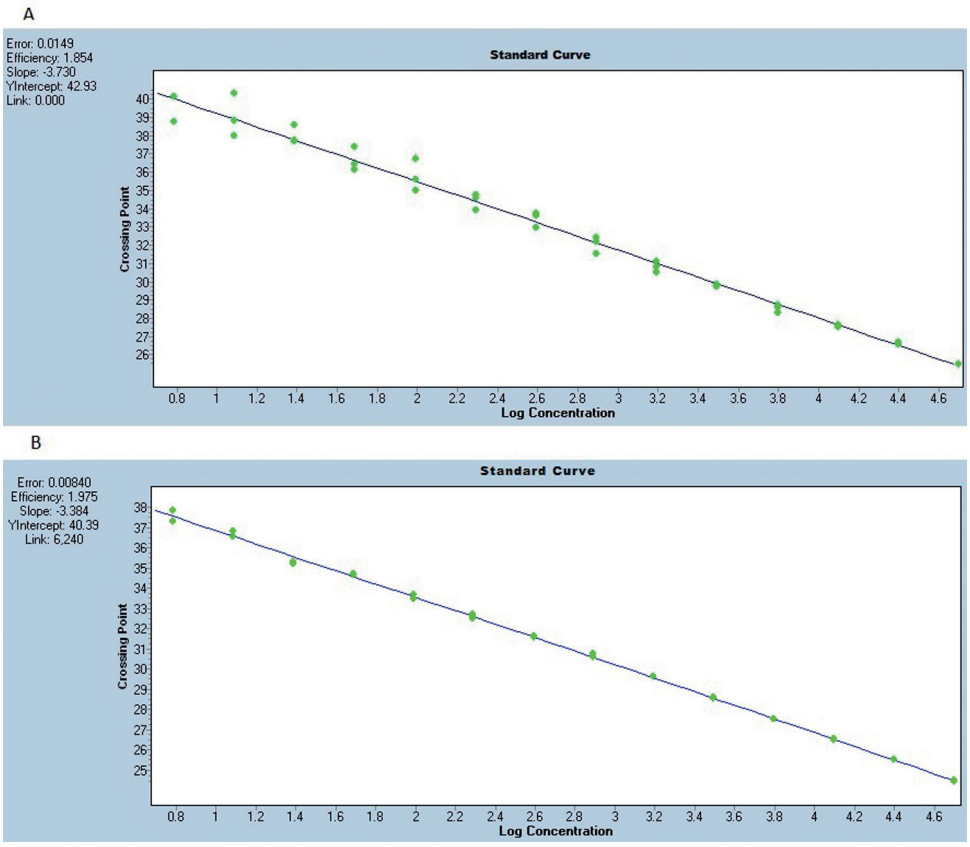


Figure S1: Standard curves based on two-fold dilutions for HPV16 and HPV18. Concentrations range from 5.0 x 10^4 to 5.0 x 10^0 copies/reaction. A: HPV16 R^2 = 0.985, B: HPV18, R^2 = 0.983.

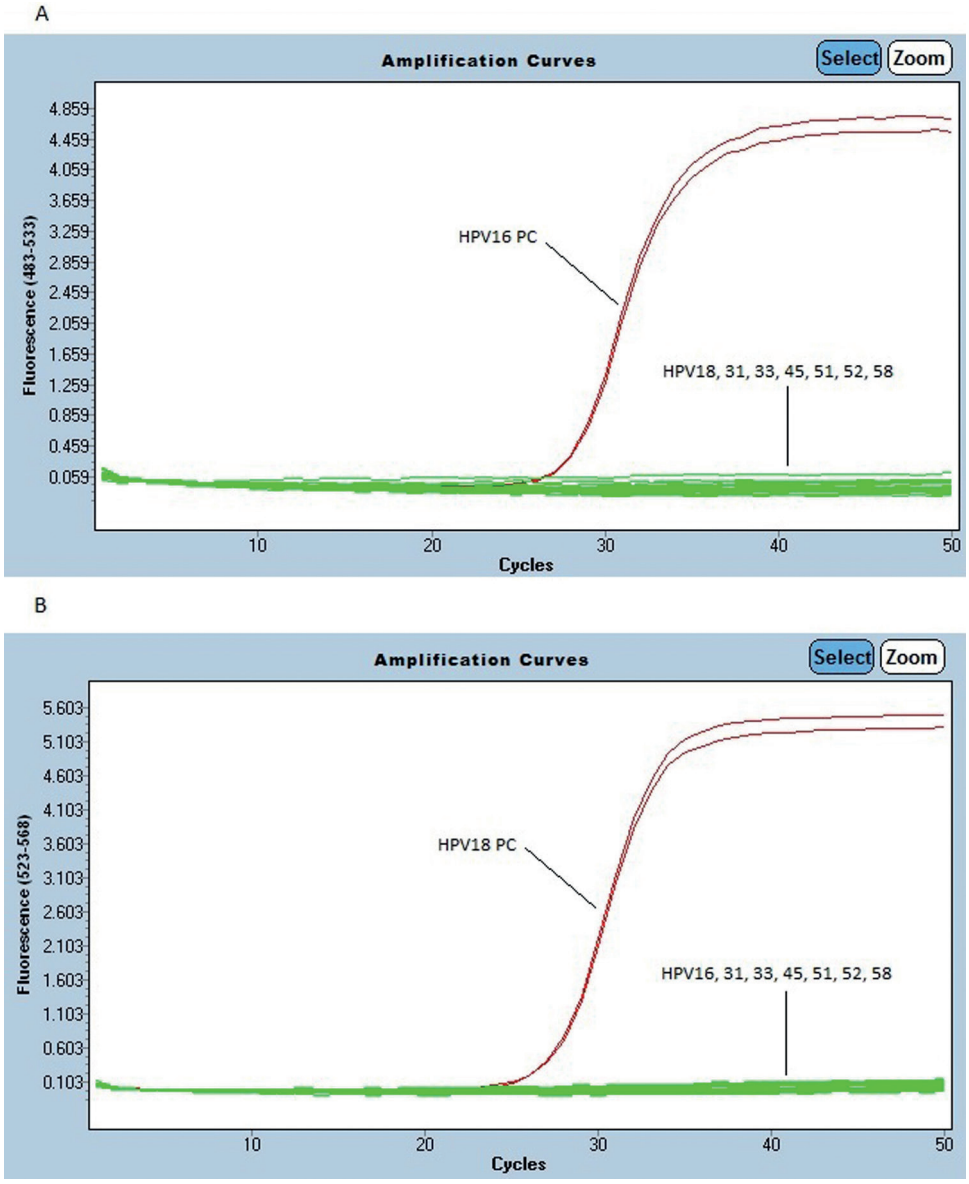
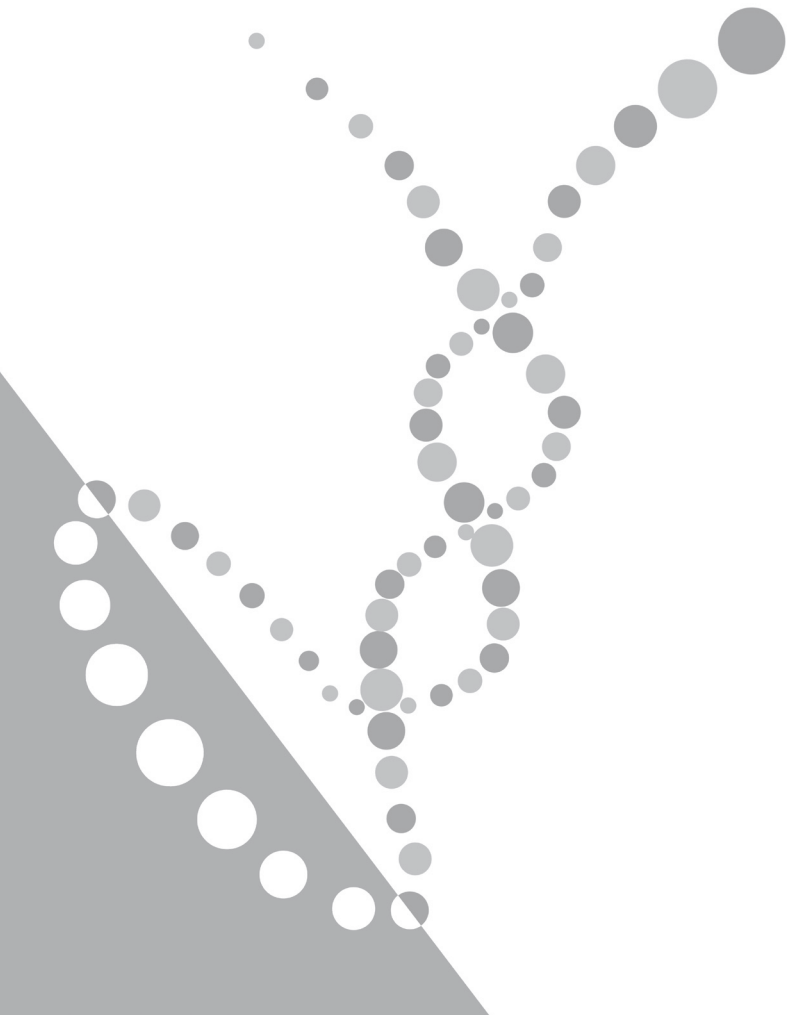


Figure S2: Specificity tests for HPV16 and HPV18 qPCR assays. A: HPV16 qPCR specificity with HPV18, 31, 33, 45, 51, 52 and 58 plasmids, PC (positive control) is HPV16 plasmid. B: HPV18 qPCR specificity with HPV16, 31, 33, 45, 51, 52 and 58 plasmids, PC is HPV18 plasmid.

Chapter 3



Effect of the bivalent HPV Vaccine on Viral Load of vaccine and non-vaccine HPV Types in incident clearing and persistent Infections in young Dutch Females

Pascal van der Weele[1,2], Martijn Breeuwsma[1], Robine Donken[1,2,#a,#b], Elske van Logchem[1], Naomi van Marm-Wattimena[1], Hester de Melker[1], Chris J.L.M. Meijer[2], Audrey J. King[1]

[1]National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven, the Netherlands
[2]Vrije Universiteit–University Medical Center (VUmc), Department of Pathology, Amsterdam, the Netherlands
[#a]Current address: University of British Columbia, BC Children's Hospital Research Institute, Vancouver BC, Canada
[#b]Current address: University of British Columbia, BC Women's Health Research Institute, Vancouver BC, Canada

PLoS One. 2019 Mar 4;14(3)

Abstract

Background HPV vaccination with the bivalent vaccine is efficacious against HPV16 and 18 infections and cross-protection against non-vaccine HPV types has been demonstrated. Here, we assessed (cross-) protective effects of the bivalent HPV16/18 vaccine on incident and persistent infections and viral load (VL) of fifteen HPV types in an observational cohort study monitoring HPV vaccine effects.

Methods Vaginal samples were obtained annually. Type-specific VL assays were developed for HPV6,11,31 33,35,39,45,51,52,56,58,59 and 66 and used in addition to existing HPV16 and 18 assays. Rate differences of incident clearing and persistent infections were correlated with differences in VL and vaccination status.

Results HPV16/18 vaccination resulted in significantly lower incidence of HPV16/18 infections and significantly lower VL in breakthrough HPV16 ($p < 0.01$) and 18 infections ($p < 0.01$). The effects of vaccination on non-vaccine type VL were ambiguous. Incidence and/or persistence rates of HPV31, 33, 35 and 45 were reduced in the vaccinated group. However, no significant type specific VL effects were found against HPV31, 33, 45, 52 in the vaccinated group. For HPV 6, 59 and 66 no significant reductions in numbers of incident and persistent infections were found, however borderline) VL reductions following vaccination were observed for HPV6 ($p = 0.01$), 59 ($p = 0.10$) and 66 ($p = 0.03$), suggesting a minor effect of the vaccine on the VL level of these HPV types. Overall, vaccination resulted in infections with slightly lower VL, irrespective of HPV type.

Conclusions In conclusion, vaccination with the bivalent HPV16/18 vaccine results in significantly reduced numbers of HPV16 and 18 incidence rates and reduced VL in breakthrough infections. Significant reductions in incident and/or persistent HPV31, 33, 35 and 45 infections were found, but no significant effect was observed on the VL for infections with these types. For the other non-vaccine HPV types no reduction in incident and/or persistent infections were found, but overall the VL tended to be somewhat lower in vaccinated women.

Introduction

Cervical intraepithelial lesions (CIN) and cervical cancer are caused by persistent high-risk human papillomavirus (hrHPV) infections [1]. Progression from initial infection to precancerous lesions and ultimately cervical cancer can take decades [2, 3]. CIN2 and CIN3 are the most advanced precursor lesions for cervical cancer and are accepted by the WHO as surrogate cervical cancer endpoint markers for vaccine efficacy. Later, persistent HPV infections (>6 months) were also accepted as adequate surrogate endpoint markers [4]. Among hrHPV types, HPV16 and 18 cause approximately 70% of all cervical cancers worldwide. A further 20% of the worldwide cervical cancer burden is added by HPV31, 33, 45, 52 and 58 [5].

To prevent HPV-associated malignancies, three highly efficacious prophylactic recombinant vaccines are currently registered and commercially available. The bivalent vaccine protects against HPV16 and 18, quadrivalent adds HPV6 and 11, and nonavalent further adds HPV31, 33, 45, 52 and 58 [6]. Varying degrees of cross-protection have been described against non-vaccine oncogenic HPV types for the bivalent and the quadrivalent vaccines [7–15]. The bivalent vaccine does seem to generate a broader cross-protective response against non-vaccine HPV types associated with CIN2+ (HPV31, 33, 39, 45 and 51) than the quadrivalent vaccine (limited to HPV31) [16]. Interestingly, limited VE has also been reported for the bivalent vaccine against HPV6 and 11 infections [17], resulting in a reduced occurrence of anogenital warts in England [18, 19]. However, these results should be further confirmed. Cross-protection is currently a heavy focus of investigation in the field and cross-protective effects are being (partly) attributed to the AS04 adjuvant (3-*O*-desacyl-4'-monophosphoryl lipid A adsorbed on aluminum salt), which appears to induce enhanced and longer-lasting antibody titers than the aluminum hydroxyphosphate sulfate adjuvant used in the quadri- and nonavalent vaccines [20]. Vaccination with the bivalent HPV vaccine also offers high, but not complete, protection against vaccine type incident infections [21, 22].

The protective effect against persistent HPV16/18 and some non-vaccine type HPV infections and associated CIN2+ lesions [11, 12, 16, 21–24], combined with the lower VE against incident HPV16/18 infections [21, 22], suggest that infections that establish in vaccinated individuals are very unlikely to persist. It has been previously described that persistence of HPV infection is associated with high viral load (VL) [25, 26]. Lower VL of these HPV types could be expected among HPV infections in vaccinated compared to non-vaccinated individuals. Consequently, longitudinal VL measurements might give additional insights in vaccine effects.

In 2009, bivalent HPV vaccination was introduced in the Netherlands. At the same time, the HPV Amongst Vaccinated and Non-vaccinated Adolescents (HAVANA) cohort study was started to monitor vaccine effectiveness [12, 27, 28]. Recently, detailed analyses of this cohort were published, including type-specific and pooled vaccine effectiveness estimates against incident and persistent infections. A high vaccine effectiveness against HPV16 and HPV18 infections was shown, as well as significant cross-protection against HPV31, 33 and 45 [12]. The present analysis aimed to use the same cohort study to assess type-specific (TS) incident and persistent HPV infections and to compare their VLs in both vaccinated and non-vaccinated women in order to see whether TS VL measurements could be used as a marker for HPV type protection.

Materials and methods

Study population and HPV DNA genotyping

Vaginal self-samples were obtained annually from participants of the "*HPV Amongst Vaccinated and Non-vaccinated Adolescents*" (HAVANA) observational cohort study, which has been described previously [28]. In short, 29,162 girls, aged 14–16 years old, eligible for catch-up vaccination, were invited for participation in 2009 and 2010. From the invitees, 6% consented to study participation. The inclusion criteria for the present analysis were girls having received either three doses of the bivalent HPV vaccine, or zero doses. Study characteristics for the present population were compared to a previous analysis of this study cohort [12]. Informed consent was obtained from study participants and both parents (if possible), or a legal representative. This study was approved by the Medical Ethics Committee of VUmc Amsterdam (2009/22). For the present analysis, a baseline and seven follow-up samples were available per study participant.

Total DNA was isolated from 200µl of self-samples using the MagNAPure96 platform (Roche Diagnostics) and eluted in 100µl elution buffer. A phocine herpesvirus-1 spike was added to each sample as an internal control for DNA isolation. Isolated DNA was genotyped using the analytically sensitive SPF10-DEIA-LiPA$_{25}$ platform [29–31].

Plasmid transformation, isolation and quantification

Ampicillin resistant plasmids for HPV6, 11, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66 were kindly provided by the HPV Reference Center, Karolinska Institutet, Sweden. Plasmids were transformed into heat-competent NEB 5-alpha *E. coli* (High Efficiency) (New England Biolabs) according to the manufacturer's protocol and the cell suspension was subsequently plated on MH agar plates containing ampicillin. Plates were incubated overnight at 37°C and single colonies were transferred to liquid LB medium containing ampicillin. Cultures were grown while shaking to $OD_{600} = 2–3$, and then centrifuged at 5000g for ten minutes. Plasmid was isolated from cell pellets using GeneJet Plasmid Midiprep kit (Thermo Scientific) according to the manufacturer's protocol. Isolated plasmid was eluted in 40µl elution buffer and quantitated in triplicate in three individual experiments to minimize assay variation using PicoGreen dsDNA quantification assay (ThermoFisher).

Assay performance and viral load quantification

New quantitative VL assays for HPV6, 11, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66 were developed in line with HPV16/18 assays described previously [26]. Primer and probe sequences are shown in S1 Table. Quantifications were performed on the LightCycler480 (Roche Diagnostics) platform. Calibration curves were generated for individual assays using triple measurements of twofold dilution series as described previously [26]. Detection limits were set at the lowest concentration detectable in the linear spectrum of the calibration curve (S1 Table). Assay specificity was determined by testing TS reagents with different plasmids at high concentrations ($> 1x10^7$ genome equivalents (GEq)/reaction). To assess possible over- and underscoring of assays, nine dilutions

from each calibration curve were tested in the presence and absence of high concentrations ($>1.7x10^5$ GEq/reaction) of the eight most prevalent hrHPV types in this study; HPV16, 18, 31, 33, 45, 51, 52 and 58. All measurements were performed in singleplex for optimal sensitivity and normalized by measuring cellular content via beta-actin qPCR [32]. Samples TS HPV DNA positive by genotyping, but without measurable VL were assigned an arbitrary VL of 10^{-5} copies/cell [33].

Infection and incidence rate definitions

To be included in the analysis of any HPV type, study participants had to be genotyping negative at baseline for that specific HPV type. HPV (baseline) positivity was assessed on a per type basis. Study participants could be included independently in multiple TS analyses. An exclusion for any one HPV type did not result in *de facto* exclusion for other HPV types.

Infection definitions are explained in Table 1. Any TS HPV measurement following a TS HPV negative baseline was considered an incident infection. Clearing infections were defined as a single HPV TS positive measurement with no new HPV TS positive measurement identified within two years (Table 1, C-E). Infection persistence was defined as at least two subsequent HPV TS positive measurements (±12 months between samples; Table 1 G-J). If a study participant fulfilled both the criteria for a clearing and a persistent infection for a single HPV type, only the first infection was included (Table 1 F, K). For VL analysis of persistent infections, the first positive samples were quantified, since load at first detection is a major determinant of HPV persistence [34].

Incidence rates (IR) were calculated for incident clearing (IRI) and persistent (IRP) infections to compare with VL measurements. Person-time for any infection was only counted when at least two subsequent measurements were available after the first HPV TS positive sample. As the exact moment of infection could not be determined, interval censoring was applied for incident clearing infections after the first TS HPV positive measurement. For persistent infections one more sample after the initial TS HPV positive sample was included after which the participant was censored. IRI and IRP were adjusted for factors associated with vaccination status, as reported previously: age, urbanization degree, ever smoked, ever used contraception, ever had sex [12]. Adjusted (a)IRI and aIRP ratios comparing vaccinated and non-vaccinated participants were calculated with 95% confidence intervals (CI) using a Poisson model. Ratios not overlapping unity were considered statistically significant.

Table 1. Infection definitions as used for this study.

	Baseline TS HPV	Per round (annual) TS HPV					Action taken
A	+						Excluded
B	-	-	-	-	-	-	HPV negative
C	-	+	-	-	-	-	Incident clearing
D	-	+	-	-	+	-	Incident clearing
E	-	+	-	0	0	0	Incident clearing
F	-	+	-	-	+	+	Incident clearing
G	-	+	-	+	-	+	Excluded
H	-	+	+	-	-	-	Incident persistent
I	-	+	+	-	+	+	Incident persistent
J	-	+	+	+	+	+	Incident persistent
K	-	+	+	-	-	+	Incident persistent
L	-	+	+	0	0	0	Incident persistent

Participants supplied samples annually. Participants were excluded from a type-specific (TS) analysis if positive at baseline for that human papillomavirus (HPV) type (A). Following a negative TS HPV baseline measurement, participants were either negative, incident clearing or incident persistent for that HPV type (B-F, H-L). Intermittent infections (G) were excluded due to uncertainty about true incidence or latent persistence. Infections meeting definitions for both clearing and persistent infections were only included based on the first infection identified for that HPV type (F, K).

Statistical analysis

Statistical analysis was performed using GraphPad Prism 7 (GraphPad Software) and SAS 9.4 (SAS Institute Inc.). Two-tailed p-values < 0.05 were considered statistically significant. Differences in study populations for the present analysis and the study population as published by Donken *et al.* [12] were assessed via chi square and student's t-test. VL differences were initially assessed using the Mann-Whitney U test based on vaccination status for the complete dataset and then stratified for incident and persistent infections.

Results

Study characteristics

From a total of 1832 HAVANA study participants, 115 women were excluded due to a missing sample at baseline or incomplete or unknown vaccination status. 1717 non-vaccinated (n = 769; 44.8%) or three-dose vaccinated (n = 948; 55.2%) girls were included in this analysis. Characteristics of the selected study population are listed in Table 2 and did not differ significantly for any of the tested variables when compared to the study population described by Donken *et al.* [12]. In total, 621 incident clearing infections (327 in vaccinated and 294 in non-vaccinated) and 484 incident persistent infections (232 in vaccinated and 252 in non-vaccinated) were identified. The most prevalent HPV types by absolute count in both vaccine and non-vaccine recipients were HPV51 (n = 123; n = 100 respectively), HPV66 (n = 88; n = 65), and HPV52 (n = 59; n = 57), as shown in Tables 3 and 4. The present analysis was compared to a previous epidemiological description of this cohort study [12] and differences were assessed.

Table 2: General characteristics at baseline for the study population subset.

	Present analysis	Donken *et al.* [12]	
Vaccination coverage			
n (%)	948 (55)	875 (54)	p = 0.32
Mean age (range)			
Non-vaccinated	15 (14–17)	15 (14–17)	p = 0.99
Vaccinated	15 (14–16)	15 (14–16)	p = 0.90
Low urbanisation			
Non-vaccinated	227 (31)	227 (31)	p = 0.95
Vaccinated	126 (14)	114 (13)	
Ever smoked			
Non-vaccinated	249 (38)	248 (38)	p = 0.89
Vaccinated	282 (32)	267 (32)	
Ever used contraception			
Non-vaccinated	314 (42)	314 (42)	p = 0.82
Vaccinated	350 (38)	332 (38)	
Ever had sex			
Non-vaccinated	210 (28)	210 (28)	p = 0.97
Vaccinated	201 (22)	187 (22)	

HPV16/18 vaccination leads to reduced incidence of HPV16/18, with some cross-protective effects

Incidence rates and ratios were calculated per thousand observed person years for vaccinated and non-vaccinated individuals and adjusted according to Donken *et al.* [12], as shown in Tables 3 and 4. In line with results reported by Donken *et al*. [12], we find significantly less incident clearing and persistent HPV16 infections in the vaccinated group (aIRI ratio: 0.38; 95%CI: 0.17–0.87; aIRP ratio: 0.08; 95%CI: 0.02–0.27). For HPV18 no significant effect was observed against clearing infections, but persistent infections did not occur at all in vaccinated individuals (no ratio calculated, vaccinated IRP of zero).

For non-vaccine types, a reduced incidence of incident clearing infections was found in the vaccinated group for HPV31 (aIRI ratio: 0.32, 95%CI: 0.13–0.76), HPV33 (aIRI ratio: 0.31, 95%CI: 0.11–0.88), HPV35 (no ratio calculated, vaccinated aIRI of zero) and HPV45 (aIRI ratio: 0.07, 95%CI: 0.01–0.56). Additionally, less HPV31 persistent infections were found among vaccinated study participants (aIRP ratio: 0.18, 95%CI: 0.06–0.52). Interestingly, more HPV59 persistent infections were found in the vaccinated group (aIRP ratio: 8.57, 95%CI: 1.06–69.04). For other HPV types, no (significant) differences were observed.

Table 3: Infection characteristics of the study cohort, stratified by vaccination status and HPV type for incident clearing infections.

	Incident clearing infections									
	Non-vaccinated				Vaccinated				IRI Ratio	
	Cases	PY	IR per 1000 PY	Adjusted	Cases	PY	IR per 1000 PY	Adjusted	Ratio (95%CI)	Adjusted
HPV6	35	3490	10.03 (7.20–13.97)	9.04 (6.20–13.17)	39	4153	9.39 (6.86–12.85)	8.24 (5.82–11.67)	0.94 (0.59–1.48)	0.91 (0.55–1.50)
HPV11	10	3625	2.76 (1.48–5.13)	2.33 (1.13–4.82)	12	4266	2.81 (1.60–4.95)	2.79 (1.53–5.08)	1.02 (0.44–2.36)	1.20 (0.47–3.02)
HPV16	22	3487	6.31 (4.15–9.58)	5.53 (3.37–9.05)	9	4279	2.10 (1.10–4.04)	2.12 (1.08–4.14)	0.33 (0.15–0.72)	**0.38 (0.17–0.87)**
HPV18	12	3567	3.37 (1.91–5.93)	1.52 (0.67–3.48)	13	4270	3.05 (1.77–5.24)	1.41 (0.64–3.08)	0.91 (0.41–1.98)	0.92 (0.39–2.21)
HPV31	22	3544	6.21 (4.09–9.43)	5.36 (3.24–8.89)	7	4285	1.63 (0.78–3.43)	1.70 (0.80–3.59)	0.26 (0.11–0.62)	**0.32 (0.13–0.76)**
HPV33	15	3604	4.16 (2.51–6.91)	3.62 (1.96–6.67)	5	4284	1.17 (0.49–2.80)	1.12 (0.45–2.77)	0.28 (0.10–0.77)	**0.31 (0.11–0.88)**
HPV35	5	3642	1.37 (0.57–3.30)	0.60 (0.11–3.20)	0	4317	-	-	-	-
HPV39	24	3544	6.78 (4.54–10.10)	5.12 (3.12–8.39)	21	4190	5.01 (3.27–7.69)	3.82 (2.31–6.31)	0.74 (0.41–1.33)	0.75 (0.39–1.41)
HPV45	13	3610	3.60 (2.09–6.20)	2.75 (1.33–5.67)	1	4320	0.23 (0.03–1.64)	0.20 (0.03–1.47)	0.06 (0.01–0.49)	**0.07 (0.01–0.56)**
HPV51	46	3343	13.76 (10.31–18.37)	10.26 (7.25–14.59)	62	3917	15.83 (12.34–20.30)	14.56 (11.07–19.15)	1.15 (0.79–1.68)	1.42 (0.93–2.16)
HPV52	29	3468	8.37 (5.81–12.03)	7.19 (4.71–10.99)	22	4151	5.30 (3.49–8.05)	4.85 (3.09–7.61)	0.63 (0.36–1.10)	0.67 (0.38–1.21)
HPV56	30	3502	8.57 (5.99–12.25)	6.85 (4.47–10.52)	36	4160	8.66 (6.24–12.00)	8.46 (5.98–11.96)	1.01 (0.62–1.64)	1.23 (0.72–2.12)
HPV58	9	3623	2.48 (1.29–4.77)	1.94 (0.88–4.30)	7	4276	1.64 (0.78–3.43)	1.36 (0.60–3.12)	0.66 (0.25–1.77)	0.70 (0.25–1.95)
HPV59	15	3627	4.13 (2.49–6.86)	3.56 (1.98–6.38)	17	4234	4.02 (2.50–6.46)	3.36 (1.96–5.78)	0.97 (0.48–1.94)	0.95 (.45–1.97)
HPV66	40	3466	11.54 (8.47–15.73)	10.97 (7.78–15.46)	43	4028	10.68 (7.72–14.39)	9.22 (6.60–12.87)	0.93 (0.60–1.42)	0.84 (0.53–1.34)

Observed person years (PY) are shown, with incidence rates (IR) for incident (IRI) and persistent (IRP) infections. The ratio represents vaccinated IR per 1000PY divided by non-vaccinated IR per 1000PY. Rates were adjusted for age, urbanization degree, ever smoked, ever used contraception and ever had sex according to [12]. Significant results are displayed in bold.
-: No value could be calculated due to zero case incidence.
#: Model did not converge due to insufficient data points.

Table 4: Infection characteristics of the study cohort, stratified by vaccination status and HPV type for incident persistent infections.

	Incident persistent infections									
	Non-vaccinated				Vaccinated				IRP Ratio	
	Cases	PY	IR per 1000 PY	Adjusted	Cases	PY	IR per 1000 PY	Adjusted	Ratio (95%CI)	Adjusted
HPV6	14	3569	3.92 (2.32–6.62)	3.57 (1.97–6.47)	13	4232	3.07 (1.78–5.29)	2.15 (1.11–4.17)	0.78 (0.37–1.67)	0.60 (0.26–1.40)
HPV11	2	3642	0.55 (0.14–2.20)	#	5	4296	1.16 (0.48–2.80)	0.29 (0.04–2.00)	2.12 (0.41–10.92)	#
HPV16	33	3588	9.21 (6.54–12.95)	7.70 (5.02–11.82)	3	4291	0.69 (0.23–2.17)	0.62 (0.20–1.97)	0.08 (0.02–0.25)	**0.08 (0.02–0.27)**
HPV18	15	3612	4.15 (2.50–6.89)	3.37 (1.76–6.46)	0	4278	-	-	-	**-**
HPV31	21	3620	5.80 (3.78–8.90)	5.38 (3.25–8.88)	5	4306	1.16 (0.48–2.79)	0.94 (0.35–2.55)	0.20 (0.08–0.53)	**0.18 (0.06–0.52)**
HPV33	5	3642	1.37 (0.57–3.30)	0.86 (0.27–2.74)	5	4308	1.16 (0.48–2.79)	0.96 (0.35–2.62)	0.85 (0.24–2.92)	1.12 (0.26–4.40)
HPV35	4	3664	1.09 (0.41–2.91)	0.55 (0.13–2.29)	3	4324	0.69 (0.22–2.15)	0.57 (0.16–2.09)	0.64 (0.14–2.84)	1.05 (0.20–5.57)
HPV39	17	3611	4.71 (2.93–7.57)	3.77 (2.12–6.71)	17	4265	3.99 (2.48–6.41)	3.55 (2.09–6.01)	0.85 (0.43–1.66)	0.94 (0.43–2.04)
HPV45	3	3643	0.82 (0.27–2.55)	0.00 (0.00–0.02)	1	4324	0.23 (0.03–1.64)	0.00 (0.00–0.01)	0.28 (0.03–2.70)	0.24 (0.03–2.35)
HPV51	54	3508	15.39 (11.79–20.10)	14.78 (11.03–19.80)	61	4111	14.84 (11.55–19.07)	14.12 (10.79–18.48)	0.96 (0.67–1.39)	0.96 (0.65–1.41)
HPV52	29	3558	8.15 (5.66–11.73)	7.12 (4.68–10.81)	37	4251	8.70 (6.31–12.01)	7.87 (5.52–11.22)	1.07 (0.66–1.74)	1.11 (0.65–1.90)
HPV56	24	3595	6.68 (4.48–9.96)	5.02 (3.06–8.23)	19	4255	4.47 (2.85–7.00)	3.65 (2.20–6.07)	0.67 (0.37–1.22)	0.73 (0.38–1.40)
HPV58	5	3650	1.37 (0.57–3.29)	0.52 (0.13–1.99)	7	4306	1.62 (0.78–3.41)	0.66 (0.20–2.20)	1.19 (0.38–3.74)	1.28 (0.37–4.44)
HPV59	2	3658	0.55 (0.14–2.19)	0.25 (0.03–1.86)	11	4276	2.57 (1.43–4.65)	2.12 (0.10–4.37)	4.71 (1.04–21.23)	**8.57 (1.06–69.04)**
HPV66	25	3582	6.98 (4.72–10.33)	6.82 (4.46–10.41)	45	4173	10.78 (8.05–14.44)	10.00 (7.26–13.77)	1.55 (0.95–2.52)	1.47 (0.87–2.48)

Observed person years (PY) are shown, with incidence rates (IR) for incident (IRI) and persistent (IRP) infections. The ratio represents vaccinated IR per 1000PY divided by non-vaccinated IR per 1000PY. Rates were adjusted for age, urbanization degree, ever smoked, ever used contraception and ever had sex according to [12]. Significant results are displayed in bold.
-: No value could be calculated due to zero case incidence.
#: Model did not converge due to insufficient data points.

Type-specific quantification of HPV viral load

HPV6, 11, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66 TS VL assays were specific in the presence of at least 10^7 copies of different HPV types. Sensitivity was < 5 GEq/reaction for each individual assay. Reproducibility of VL measurements was assessed by measuring duplicates of a subset of HPV16 (n = 62) and HPV18 (n = 22) samples. Duplicates were found to be on average < 0.1 ct values apart (median 0.03; minimum -2.39; maximum 1.66). VL measurements were performed based on genotyping results. Figs 1–3 show scatter plots of VL measurements for study participants in all infections, the clearing infections subset, and persistent infections subset respectively.

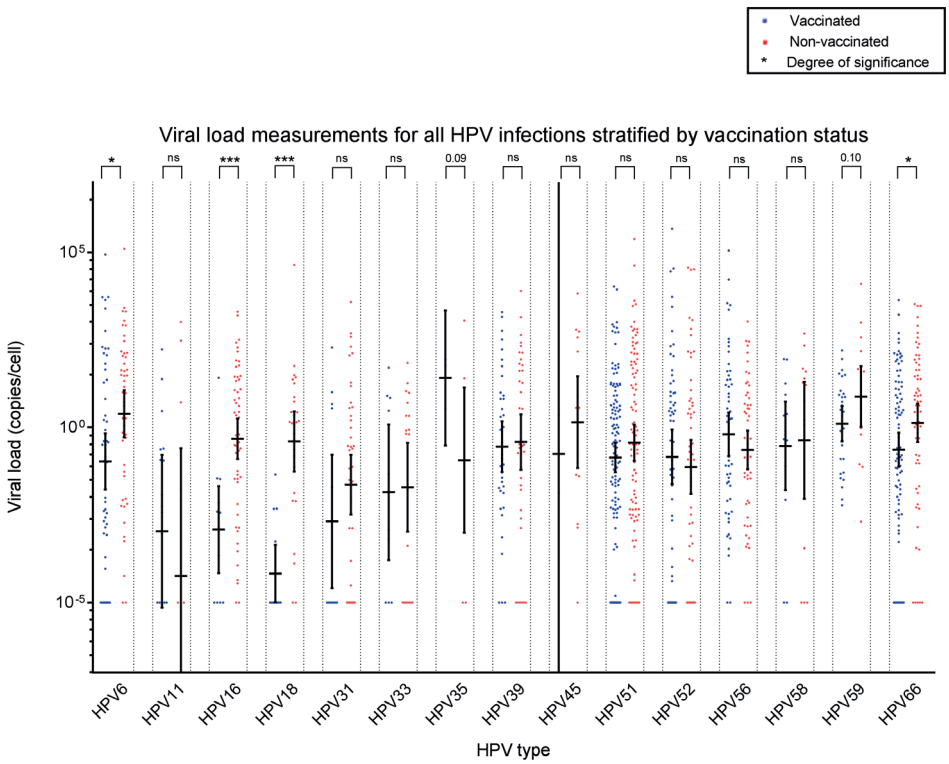


Fig 1. Viral load measurements of all infections included in this study, expressed in copies per cell on a log10 scale. Individual measurements are displayed as blue and red dots for vaccinated and non-vaccinated study participants respectively. For each HPV type median VL values were compared between vaccinated and non-vaccinated individuals. Statistical significance is shown above each HPV type, with * meaning $p < 0.05$; ** meaning $p < 0.01$; *** meaning $p < 0.001$.

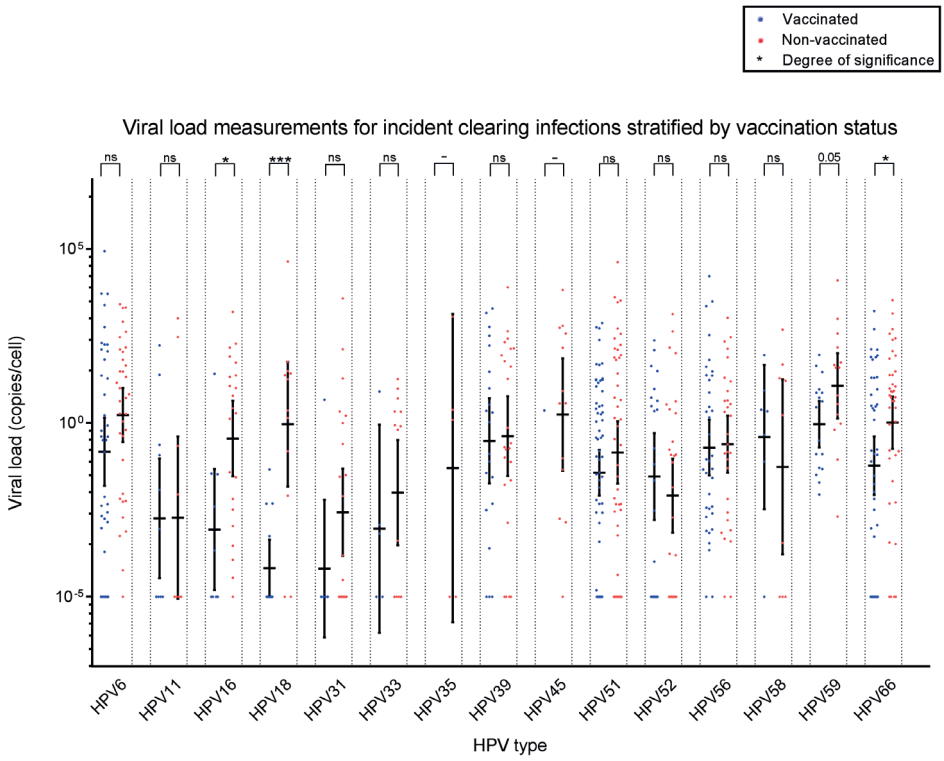


Fig 2. Viral load measurements in incident infections, expressed in copies per cell on a log10 scale. Individual measurements are displayed as blue and red dots for vaccinated and non-vaccinated study participants respectively. For each HPV type median VL values were compared between vaccinated and non-vaccinated individuals. Statistical significance is shown above each HPV type, with * meaning $p < 0.05$; ** meaning $p < 0.01$; *** meaning $p < 0.001$.

Reduced viral load in breakthrough HPV16/18 infections following HPV16/18 vaccination

VL measurements correlated well with observed reductions in incident/persistent HPV16/18 infections. Median HPV16 VL was significantly lower in vaccine recipients for the total study analysis ($p < 0.01$, Fig 1), as well as stratified analyses for incident clearing ($p = 0.02$, Fig 2) and persistent infections ($p = 0.05$, Fig 3). For HPV18, vaccine recipients showed significantly lower median VL in the total analysis ($p < 0.01$, Fig 1) and in the incident clearing analysis ($p < 0.01$, Fig 2). Since no persistent HPV18 infections were identified among vaccinees, no VL comparison could be performed.

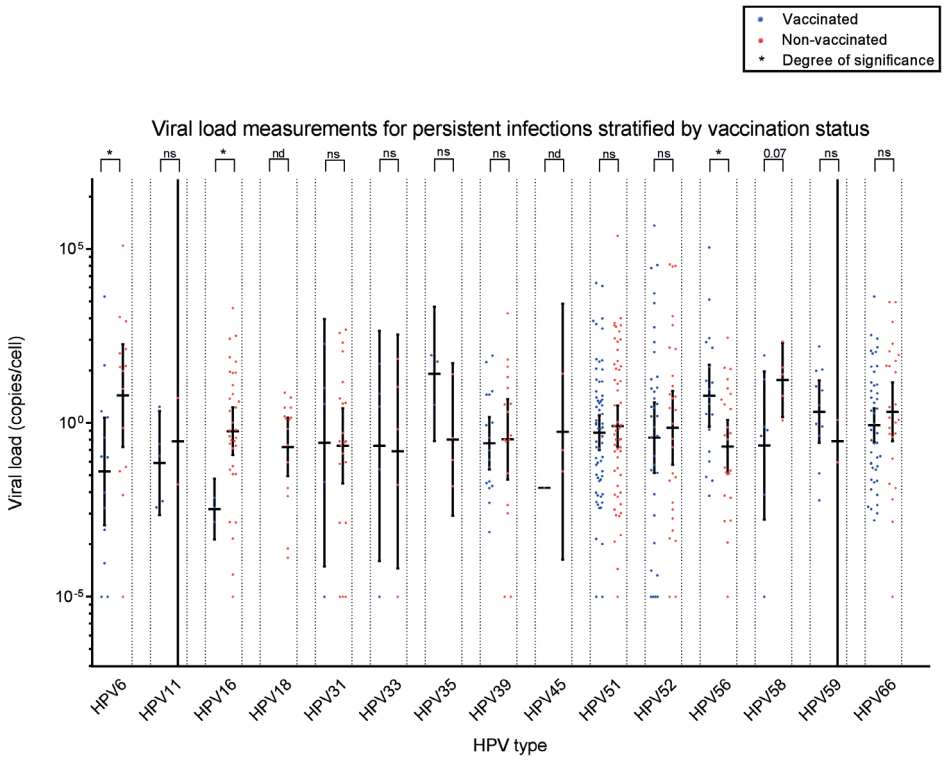


Fig 3. Viral load measurements in persistent infections, expressed in copies per cell on a log10 scale. Individual measurements are displayed as blue and red dots for vaccinated and non-vaccinated study participants respectively. For each HPV type median VL values were compared between vaccinated and non-vaccinated individuals. Statistical significance is shown above each HPV type, with * meaning $p < 0.05$; ** meaning $p < 0.01$; *** meaning $p < 0.001$.

Cross-protective effects on viral load following vaccination are unclear

Despite finding some cross-protection of the bivalent vaccine against HPV31, 33, 35 and 45 (Tables 3 and 4), no significant differences in median VL were observed between groups in any of the analyses for these types. Except for a marginally increased median HPV35 VL in the vaccinated group for the total analysis ($p = 0.09$), although this was based on a limited number of infections ($n = 3$ vaccinated, $n = 9$ non-vaccinated). For HPV6, significantly lower median VL values were found in the vaccinated group for the total ($p = 0.01$) and the persistent infections analyses ($p = 0.05$). For HPV59 and 66 (borderline) lower median VL values were found in the vaccinated

group for the total (p = 0.10; p = 0.03) and incident clearing (p = 0.05; p = 0.04) infection analyses. For HPV58 a marginally lower median VL was observed in vaccinated individuals (p = 0.07). For HPV56 an increased median VL was observed in the vaccinated group (p = 0.03). The VL results for HPV6, 56, 58, 59 and 66 did not lead to a measurable decrease in the rate of incident or persistent infections in the vaccinated group. Interestingly, the median VL of any HPV type in the vaccinated group shows a trend in being slightly lower than in the non-vaccinated group (Figs 1–3). To indicate compatibility between incidence and VL analyses, Table 5 shows summarized results per HPV type, which shows good concordance between analyses for vaccine types, but poor concordance for non-vaccine types.

Table 5: Summary and compatibility of HPV infection incidence with viral load.

	Vaccinated compared to non-vaccinated groups; effects measured on:				
	Infection incidence rate			**Viral load**	
HPV type	**Incident clearing**	**Incident persistent**	**Total infections**	**Clearing infections**	**Persistent infections**
6	-	-	Reduced	-	Reduced
11	-	-	-	-	-
16	Reduced	Reduced	Reduced	Reduced	Reduced
18	-	Reduced [1]	Reduced	Reduced	- [1]
31	Reduced	Reduced	-	-	-
33	Reduced	-	-	-	-
35	Reduced [1]	-	Increased [2]	- [1]	-
39	-	-	Reduced [2]	-	-
45	Reduced	-	-	- [1]	-
51	-	-	-	-	-
52	-	-	-	-	-
56	-	-	-	-	Increased
58	-	-	-	-	Reduced [2]
59	-	Increased	Reduced [2]	Reduced	-
66	-	-	Reduced	Reduced	-

Reduced/increased incidence or viral load implies a reduction in the vaccinated group compared to the non-vaccinated group.
[1]: Calculation for p-value could not be performed due to complete absence of infections in the vaccinated group.
[2]: Borderline significant results ($0.05 < p < 0.1$).

Discussion

In this study, newly developed and existing assays [26] were used to measure VL of HPV6, 11, 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66 infections on a prospective longitudinal observational cohort study of young women who were eligible for HPV vaccination according to a three-dose schedule. To our knowledge, we are the first to study effects of HPV vaccination on the VL of HPV infections. We confirm that vaccination with the bivalent vaccine leads to significant reductions in the incidence of clearing and/or persistence of HPV16/18 infections, in line with previous results for a different sample subset of this study by Donken *et al.* [12]. In addition, we show that vaccination is associated with significant reductions in median VL of breakthrough incident clearing and persistent HPV16/18 infections. Vaccination with the bivalent vaccine also induced a reduction in both incident clearing and persistent HPV31 infections as well as reductions in incident clearing HPV33 and 45 infections. However, these findings were not associated with any reductions in VL of these HPV types following vaccination. Our findings suggest that vaccination is significantly associated with reduction in VL in breakthrough vaccine type infections, possibly by limiting the capacity of the virus to cause a persistent infection. However, despite confirming cross-protection against HPV31, 33 and 45, the effect on VL is not significantly altered, suggesting that cross-protection against non-vaccine HPV types is either complete (preventing infections in a similar fashion to vaccine types) or absent (not measurable by VL in this study).

Cross-protection has been previously described for HPV31, 33, 35, 45, 52 and 58 [7–10]. We could confirm some of these effects in this study (HPV31, 33, 35 and 45), but no significant VL reductions in the vaccinated group were found for either type, despite a reduced effect , although HPV33 and 45 infections are uncommon in this study (n = 10 for HPV33, n = 4 for HPV45). For HPV6, 59 and 66 (borderline) significant median VL reductions were found among vaccine recipients, although this did not translate into any measurable reduction of incident clearing or persistent infections for these HPV types in this study. Cross-protection is less likely against HPV6 and 66, as they belong to a different HPV species than HPV16 or 18 [35]. For HPV59, despite it being closely related to HPV18, the SPF10-DEIA-LiPA$_{25}$ genotyping assay has a relatively poor detection sensitivity [36], which might explain the lack of concordance between HPV59 VL and incidence results.

Vaccination (with any of the currently available virus-like particle (VLP) based vaccines) primarily leads to sterilizing immunity against HPV16/18 via type-restricted neutralizing antibodies, preventing the majority of incident and persistent infections [37]. We report significantly lower median VL values in vaccinated individuals with HPV16/18 breakthrough infections. Our results could suggest that the immune response protecting against HPV infection acquisition could be different from the immune response responsible for viral control. This effect might be facilitated by a T cell mediated immune response limiting viral replication, which is the proposed mechanism leading to cross-protection against non-vaccine type associated CIN2+ [37, 38], and which could help explain results described by Harper *et al.* [16]. This could result in infections with lower VL levels, which have been associated with clearance of infection [39, 40], while high HPV16/18 VL levels have been associated with persisting infections [25, 26].

While we are unable to clearly associate vaccine effects on incidence of infections with VL, we do observe a general trend where median VL for any HPV type was lower for vaccine recipients than non-vaccinated individuals. If these findings are indeed caused by the vaccine, they might help alleviate or prevent long-term disease effects by dampening infection development through reduced viral load.

Several limitations need to be addressed for this study. First, we included participants based on baseline TS HPV DNA negative samples, since serology was only available for a subset of our cohort. While baseline DNA positivity is informative for the present sample, it does not inform about any past exposure to HPV and any possible effects on VL. Fortunately, due to low sexual activity at inclusion for this study (Table 2), we believe prior infections should be of limited impact for the present analysis. Second, our categorical definitions of incident clearing and persistent infections (Table 1), though pragmatic and in line with previous literature use, might not accurately represent a natural situation and do not take possible deposition of HPV into account [41]. Incident persistent infections should suffer less from HPV depositions, as repeated detection of deposition events one year apart seems unlikely. Based on a previous study describing TS clearance windows for individual HPV types, one year for persistent infections is adequate in general [42]. Ideally, longer intervals should be observed before infections can be truly considered persistent. However, within this study of young, vaccinated women, infections are relatively scarce and longer interval persistent infections would lead to sharp reductions in incidence, prohibiting statistical interpretation of results. Our definitions also lead to exclusion of intermittent infections. For intermittence, it is unclear whether the observed infection is actually persistent with potential latency in between, or actual a repeated incidence of the same genotype. Since, previous research has shown that relative risk for repeat infections might be different from initial acquisition, these infections were excluded [43]. Third, VL is a highly heterogeneous parameter [26, 44]. Combined with a relatively low number of total infections in this study, this could explain why VL point estimates (Figs 1–3) are often lower in the vaccinated group than in the non-vaccinated group; yet no statistical significance is reached in comparisons and compatibility between incidence rates and VL results is low (Table 5), except for vaccine types HPV16/18. Finally, during VL measurements, a number of samples remained VL negative, despite positive genotyping results. For these samples, an artificial VL was assigned [33]. This method was chosen over assigning a VL at the detection limit of the HPV type tested, since this would lead to exaggerated results when correcting for cellular content.

Conclusions

Combined, our results suggest that vaccination against HPV16/18 affects VL in breakthrough infections when compared to non-vaccinated women. Although some effects are observed against non-vaccine types, this study is insufficiently powered to suggest a clear correlation between vaccination and effects on VL of non-vaccine HPV types. Further population-based studies are required to identify which effects are truly causal to and maintained by the vaccine.

References

1. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. The Journal of pathology. 1999;189(1):12–9. doi: 10.1002/(SICI)1096–9896(199909)189:1 < 12::AID-PATH431 > 3.0.CO;2-F. PubMed PMID: 10451482.
2. Vink MA, Bogaards JA, van Kemenade FJ, de Melker HE, Meijer CJ, Berkhof J. Clinical progression of high-grade cervical intraepithelial neoplasia: estimating the time to preclinical cervical cancer from doubly censored national registry data. American journal of epidemiology. 2013;178(7):1161–9. doi: 10.1093/aje/kwt077. PubMed PMID: 23897645.
3. Ho GY, Bierman R, Beardsley L, Chang CJ, Burk RD. Natural history of cervicovaginal papillomavirus infection in young women. The New England journal of medicine. 1998;338(7):423–8. doi: 10.1056/NEJM199802123380703. PubMed PMID: 9459645.
4. IARC HPV Working Group. Primary End-points for Prophylactic HPV Vaccine Trials. Lyon (FR)2014.
5. de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. The Lancet Oncology. 2010;11(11):1048–56. doi: 10.1016/S1470–2045(10)70230–8. PubMed PMID: 20952254.
6. Joura EA, Giuliano AR, Iversen OE, Bouchard C, Mao C, Mehlsen J, et al. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. The New England journal of medicine. 2015;372(8):711–23. doi: 10.1056/NEJMoa1405044. PubMed PMID: 25693011.
7. Scherpenisse M, Schepp RM, Mollers M, Meijer CJ, Berbers GA, van der Klis FR. Characteristics of HPV-specific antibody responses induced by infection and vaccination: cross-reactivity, neutralizing activity, avidity and IgG subclasses. PloS one. 2013;8(9):e74797. doi: 10.1371/journal.pone.0074797. PubMed PMID: 24058629; PubMed Central PMCID: PMC3776846.
8. Malagon T, Drolet M, Boily MC, Franco EL, Jit M, Brisson J, et al. Cross-protective efficacy of two human papillomavirus vaccines: a systematic review and meta-analysis. The Lancet infectious diseases. 2012;12(10):781–9. doi: 10.1016/S1473–3099(12)70187–1. PubMed PMID: 22920953.
9. Harari A, Chen Z, Rodriguez AC, Hildesheim A, Porras C, Herrero R, et al. Cross-protection of the Bivalent Human Papillomavirus (HPV) Vaccine Against Variants of Genetically Related High-Risk HPV Infections. The Journal of infectious diseases. 2016;213(6):939–47. doi: 10.1093/infdis/jiv519. PubMed PMID: 26518044; PubMed Central PMCID: PMCPMC4760417.
10. Wheeler CM, Castellsague X, Garland SM, Szarewski A, Paavonen J, Naud P, et al. Cross-protective efficacy of HPV-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by non-vaccine oncogenic HPV types: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. The Lancet Oncology. 2012;13(1):100–10. doi: 10.1016/S1470–2045(11)70287-X. PubMed PMID: 22075170.
11. Pollock KGJ. Sustained cross-protection of the bivalent HPV vaccine. The Journal of infectious diseases. 2018;217(10):1515–6. doi: 10.1093/infdis/jiy068. PubMed PMID: 29408966.
12. Donken R, King AJ, Bogaards JA, Woestenberg PJ, Meijer C, de Melker HE. High effectiveness of the bivalent HPV vaccine up to six years post-vaccination against incident and persistent HPV infections in young Dutch females. The Journal of infectious diseases. 2018;217(10):1579–89. doi: 10.1093/infdis/jiy067. PubMed PMID: 29409034.

13. Woestenberg PJ, King AJ, van Benthem BHB, Donken R, Leussink S, van der Klis FRM, et al. Bivalent Vaccine Effectiveness Against Type-Specific HPV Positivity: Evidence for Cross-Protection Against Oncogenic Types Among Dutch STI Clinic Visitors. The Journal of infectious diseases. 2018;217(2):213–22. doi: 10.1093/infdis/jix582. PubMed PMID: 29140439.
14. Brown DR, Kjaer SK, Sigurdsson K, Iversen OE, Hernandez-Avila M, Wheeler CM, et al. The impact of quadrivalent human papillomavirus (HPV; types 6, 11, 16, and 18) L1 virus-like particle vaccine on infection and disease due to oncogenic nonvaccine HPV types in generally HPV-naive women aged 16–26 years. The Journal of infectious diseases. 2009;199(7):926–35. doi: 10.1086/597307. PubMed PMID: 19236279.
15. Wheeler CM, Kjaer SK, Sigurdsson K, Iversen OE, Hernandez-Avila M, Perez G, et al. The impact of quadrivalent human papillomavirus (HPV; types 6, 11, 16, and 18) L1 virus-like particle vaccine on infection and disease due to oncogenic nonvaccine HPV types in sexually active women aged 16–26 years. The Journal of infectious diseases. 2009;199(7):936–44. doi: 10.1086/597309. PubMed PMID: 19236277.
16. Harper DM, DeMars LR. HPV vaccines–A review of the first decade. Gynecologic oncology. 2017;146(1):196–204. doi: 10.1016/j.ygyno.2017.04.004. PubMed PMID: 28442134.
17. Szarewski A, Skinner SR, Garland SM, Romanowski B, Schwarz TF, Apter D, et al. Efficacy of the HPV-16/18 AS04-adjuvanted vaccine against low-risk HPV types (PATRICIA randomized trial): an unexpected observation. The Journal of infectious diseases. 2013;208(9):1391–6. doi: 10.1093/infdis/jit360. PubMed PMID: 24092907; PubMed Central PMCID: PMCPMC3789574.
18. Howell-Jones R, Soldan K, Wetten S, Mesher D, Williams T, Gill ON, et al. Declining genital Warts in young women in england associated with HPV 16/18 vaccination: an ecological study. The Journal of infectious diseases. 2013;208(9):1397–403. doi: 10.1093/infdis/jit361. PubMed PMID: 24092908; PubMed Central PMCID: PMCPMC3789575.
19. Canvin M, Sinka K, Hughes G, Mesher D. Decline in genital warts diagnoses among young women and young men since the introduction of the bivalent HPV (16/18) vaccination programme in England: an ecological analysis. Sexually transmitted infections. 2016;93(2):125–8. doi: 10.1136/sextrans-2016–052626. PubMed PMID: 27365492.
20. Einstein MH, Levin MJ, Chatterjee A, Chakhtoura N, Takacs P, Catteau G, et al. Comparative humoral and cellular immunogenicity and safety of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine and HPV-6/11/16/18 vaccine in healthy women aged 18–45 years: follow-up through Month 48 in a Phase III randomized study. Human vaccines & immunotherapeutics. 2014;10(12):3455–65. doi: 10.4161/hv.36117. PubMed PMID: 25483700; PubMed Central PMCID: PMCPMC4514093.
21. Harper DM. Impact of vaccination with Cervarix (trade mark) on subsequent HPV-16/18 infection and cervical disease in women 15–25 years of age. Gynecologic oncology. 2008;110(3 Suppl 1):S11–7. doi: 10.1016/j.ygyno.2008.06.029. PubMed PMID: 18649932.
22. Naud PS, Roteli-Martins CM, De Carvalho NS, Teixeira JC, de Borba PC, Sanchez N, et al. Sustained efficacy, immunogenicity, and safety of the HPV-16/18 AS04-adjuvanted vaccine: final analysis of a long-term follow-up study up to 9.4 years post-vaccination. Human vaccines & immunotherapeutics. 2014;10(8):2147–62. doi: 10.4161/hv.29532. PubMed PMID: 25424918.

23. Apter D, Wheeler CM, Paavonen J, Castellsague X, Garland SM, Skinner SR, et al. Efficacy of human papillomavirus 16 and 18 (HPV-16/18) AS04-adjuvanted vaccine against cervical infection and precancer in young women: final event-driven analysis of the randomized, double-blind PATRICIA trial. Clinical and vaccine immunology : CVI. 2015;22(4):361–73. doi: 10.1128/CVI.00591–14. PubMed PMID: 25651922; PubMed Central PMCID: PMC4375348.
24. Struyf F, Colau B, Wheeler CM, Naud P, Garland S, Quint W, et al. Post hoc analysis of the PATRICIA randomized trial of the efficacy of human papillomavirus type 16 (HPV-16)/HPV-18 AS04-adjuvanted vaccine against incident and persistent infection with nonvaccine oncogenic HPV types using an alternative multiplex type-specific PCR assay for HPV DNA. Clinical and vaccine immunology : CVI. 2015;22(2):235–44. doi: 10.1128/CVI.00457–14. PubMed PMID: 25540273; PubMed Central PMCID: PMCPMC4308870.
25. Winer RL, Xi LF, Shen Z, Stern JE, Newman L, Feng Q, et al. Viral load and short-term natural history of type-specific oncogenic human papillomavirus infections in a high-risk cohort of midadult women. International journal of cancer Journal international du cancer. 2014;134(8):1889–98. doi: 10.1002/ijc.28509. PubMed PMID: 24136492; PubMed Central PMCID: PMC3947429.
26. van der Weele P, van Logchem E, Wolffs P, van den Broek I, Feltkamp M, de Melker H, et al. Correlation between viral load, multiplicity of infection, and persistence of HPV16 and HPV18 infection in a Dutch cohort of young women. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology. 2016;83:6–11. doi: 10.1016/j.jcv.2016.07.020. PubMed PMID: 27522635.
27. Mollers M, King AJ, Knol MJ, Scherpenisse M, Meijer CJ, van der Klis FR, et al. Effectiveness of human papillomavirus vaccine against incident and persistent infections among young girls: Results from a longitudinal Dutch cohort study. Vaccine. 2015;33(23):2678–83. doi: 10.1016/j.vaccine.2015.04.016. PubMed PMID: 25887090.
28. Mollers M, Scherpenisse M, van der Klis FR, King AJ, van Rossum TG, van Logchem EM, et al. Prevalence of genital HPV infections and HPV serology in adolescent girls, prior to vaccination. Cancer Epidemiol. 2012;36(6):519–24. doi: 10.1016/j.canep.2012.07.006. PubMed PMID: 22906483.
29. Kleter B, van Doorn LJ, Schrauwen L, Molijn A, Sastrowijoto S, ter Schegget J, et al. Development and clinical evaluation of a highly sensitive PCR-reverse hybridization line probe assay for detection and identification of anogenital human papillomavirus. Journal of clinical microbiology. 1999;37(8):2508–17. PubMed PMID: 10405393; PubMed Central PMCID: PMC85270.
30. Kleter B, van Doorn LJ, ter Schegget J, Schrauwen L, van Krimpen K, Burger M, et al. Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. The American journal of pathology. 1998;153(6):1731–9. doi: 10.1016/S0002–9440(10)65688-X. PubMed PMID: 9846964; PubMed Central PMCID: PMC1866345.
31. van Doorn LJ, Molijn A, Kleter B, Quint W, Colau B. Highly effective detection of human papillomavirus 16 and 18 DNA by a testing algorithm combining broad-spectrum and type-specific PCR. Journal of clinical microbiology. 2006;44(9):3292–8. doi: 10.1128/JCM.00539–06. PubMed PMID: 16954263; PubMed Central PMCID: PMC1594708.
32. Zhao M, Rosenbaum E, Carvalho AL, Koch W, Jiang W, Sidransky D, et al. Feasibility of quantitative PCR-based saliva rinse screening of HPV for head and neck cancer. International journal of cancer Journal international du cancer. 2005;117(4):605–10. doi: 10.1002/ijc.21216. PubMed PMID: 15929076.

33. Chaturvedi AK, Graubard BI, Pickard RK, Xiao W, Gillison ML. High-risk oral human papillomavirus load in the US population, National Health and Nutrition Examination Survey 2009–2010. The Journal of infectious diseases. 2014;210(3):441–7. doi: 10.1093/infdis/jiu116. PubMed PMID: 24625808; PubMed Central PMCID: PMCPMC4110460.
34. de Sanjose S, Brotons M, Pavon MA. The natural history of human papillomavirus infection. Best Pract Res Clin Obstet Gynaecol. 2018;47:2–13. doi: 10.1016/j.bpobgyn.2017.08.015. PubMed PMID: 28964706.
35. de Villiers EM. Cross-roads in the classification of papillomaviruses. Virology. 2013;445(1–2):2–10. doi: 10.1016/j.virol.2013.04.023. PubMed PMID: 23683837.
36. DDL Diagnostics Laboratory. DNA ELISA kit HPV SPF10, version 1.
37. Schiller J, Lowy D. Explanations for the high potency of HPV prophylactic vaccines. Vaccine. 2018. doi: 10.1016/j.vaccine.2017.12.079. PubMed PMID: 29325819.
38. Pinto LA, Viscidi R, Harro CD, Kemp TJ, Garcia-Pineres AJ, Trivett M, et al. Cellular immune responses to HPV-18, -31, and -53 in healthy volunteers immunized with recombinant HPV-16 L1 virus-like particles. Virology. 2006;353(2):451–62. doi: 10.1016/j.virol.2006.06.021. PubMed PMID: 16863657.
39. Carcopino X, Henry M, Mancini J, Giusiano S, Boubli L, Olive D, et al. Two years outcome of women infected with high risk HPV having normal colposcopy following low-grade or equivocal cytological abnormalities: are HPV16 and 18 viral load clinically useful predictive markers? Journal of medical virology. 2012;84(6):964–72. doi: 10.1002/jmv.23276. PubMed PMID: 22499020.
40. Trevisan A, Schlecht NF, Ramanakumar AV, Villa LL, Franco EL, Ludwig-McGill Study G. Human papillomavirus type 16 viral load measurement as a predictor of infection clearance. The Journal of general virology. 2013;94(Pt 8):1850–7. doi: 10.1099/vir.0.051722–0. PubMed PMID: 23677791.
41. Malagon T, Burchell AN, El-Zein M, Guenoun J, Tellier PP, Coutlee F, et al. Estimating HPV DNA Deposition Between Sexual Partners Using HPV Concordance, Y Chromosome DNA Detection, and Self-reported Sexual Behaviors. The Journal of infectious diseases. 2017;216(10):1210–8. doi: 10.1093/infdis/jix477. PubMed PMID: 28968731; PubMed Central PMCID: PMCPMC5853413.
42. Rositch AF, Koshiol J, Hudgens MG, Razzaghi H, Backes DM, Pimenta JM, et al. Patterns of persistent genital human papillomavirus infection among women worldwide: a literature review and meta-analysis. International journal of cancer Journal international du cancer. 2013;133(6):1271–85. doi: 10.1002/ijc.27828. PubMed PMID: 22961444; PubMed Central PMCID: PMCPMC3707974.
43. Donken R, Knol MJ, Ogilvie GS, Dobson S, King AJ, Singer J, et al. Measuring vaccine effectiveness against persistent HPV infections in observational studies: methodological challenges. Monitoring the HPV vaccination program in the Netherlands: Effects, changing schedule and future perspective 2018. p. 206–22.
44. Snijders PJ, Hogewoning CJ, Hesselink AT, Berkhof J, Voorhorst FJ, Bleeker MC, et al. Determination of viral load thresholds in cervical scrapings to rule out CIN 3 in HPV 16, 18, 31 and 33-positive women with normal cytology. International journal of cancer Journal international du cancer. 2006;119(5):1102–7. doi: 10.1002/ijc.21956. PubMed PMID: 16570279.

Supporting information

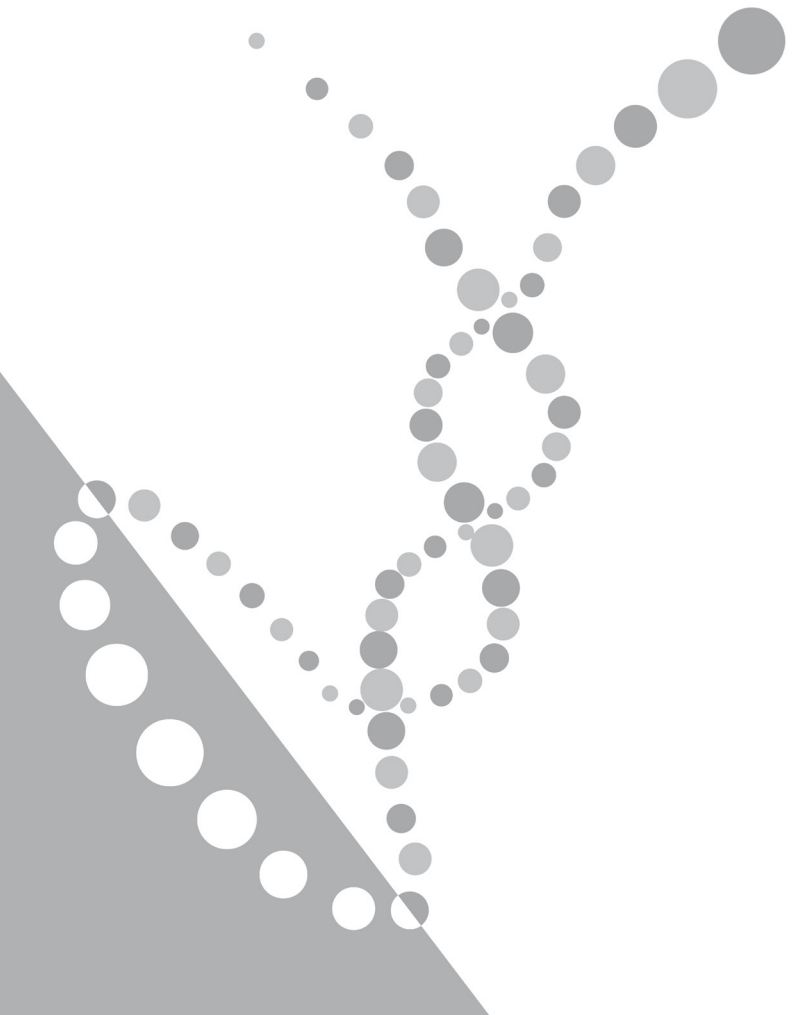
S1 Table. Sequences of primers and probes used for the quantification of different HPV types and limits of detection (LOD) of respective assays.

Genotype	Forward (5'-3')	Reverse (5'-3')	Fluoro-phore (5')	Probe (5'-3')	Quencher (3')	LOD Copies/ reaction
HPV6	CAATGGTATTTGTTGGGGTAATCAAC	TGTGGAAGATGTAGTTACGGATGCACA	FAM	ACCACACGCAGTACCAACATGACCA	EDQ	1.8
HPV11	AGGTAGGCAGGGGTCAACCT	CACAGCGTTTAGTATGGGCGTGCACA	FAM	CCAGGATTACCACCATACCCACCA	EDQ	2.6
HPV16	TTGTTGGGGTAACCAACTATTTGTTACTGTT	CCTCCCCATGTCTGAGGTACTCCTTAAAG	FAM	TGTCATTATGTGCTGCCATATCTACTTC	TAMRA	4.6
HPV18	GCATAATCAATTATTTGTTACTGTGGTAGATACCACT	GCTATACTGCTTAAATTTGGTAGCATCATATTGC	Yakima Yellow	AACAATATGTGCTTCTACACAGTCTCCTGT	BHQ-2	4.3
HPV31	GGTAGATACCACACGTAGTACCAATATGTCTGTTTG	AAATTGTAAATCAAATTCCTCACCATGTCTTA-AATACTCT	FAM	TGCTGCAATTGCAAACAGTGATACTACATT	EDQ	1.2
HPV33	GGAACTACTGCCTCTATTCAAAGCAGTGC	GTACTGTCACTAGTTACTTGTGTGCATAAAGTCATA	FAM	CTGTGGTAGATACCACTCGCAGTACTAAT	EDQ	1.0
HPV35	AAAAGGCACACCTTGTAATGCTAACCAG	AATCCATTGCACCAAATCCTGTGTC	FAM	TGTACTACAAGACGGGGACATGGT	EDQ	1.2
HPV39	CGTGGTGGTATGGTGGGTGACGCCAT	ATGGAACCGCTGGGAGAGGGGCAGTATACA	FAM	CGTGCAAACCCCGGTAGTTC	EDQ	1.4
HPV45	TAGTGGACACTACCCGCAGTACTAA	CCACATGTCTACTATACTGCTTAAACTTAGTAG	FAM	ACACAAAATCCTGTGCCAAGTACATATGAC	EDQ	1.8
HPV51	AACAATCAGCTTTTTATTACCTGTGTTGATACTACCA	CTTAAAGTTACTTGGAGTAAATGTTGGGGAAACCG	FAM	CAAATTTAACTATTAGCACTGCCACTGCT	EDQ	1.5
HPV52	TTGGGGCAATCAGTTGTTTGTCACAGTT	CGCCATGACGAAGGTATTCCTTAAAATTTTCATTTT	FAM	TTATGTGCTGAGGTTAAAAAGGAAAG-CACA	EDQ	1.8
HPV56	CCCGCTATGGGTGAACATTGGA	AATGCAAGAGGCGGGCAGTC	FAM	AGGTGCTGTGTGTAAGTCCACACAA	EDQ	1.2
HPV58	CAGTTATTTGTTACCGTGGTTGATACCACTCG	TGTAAGTCATATTCTTCAACATGACGTACATATTCC	FAM	TGCACTGAAGTAACTAAGGAAGGTACA	EDQ	1.6
HPV59	CCACTATTTTGGAGGATTGGAATTTTGG	GCGGTGCGGTGTCCTTTTGA	FAM	CCGTTTTGTTCAATCTGCTGCTGT	EDQ	1.2
HPV66	GCACATTAACTAAATATGATGCCCGTGAA	AAGTTGAAACACAAACTGTAGTTCATATTCCTCCA	FAM	ATCAATCAATACCTTCGCCATG	EDQ	1.4

Part II

Development and application of (next-generation) sequencing assays in epidemiological and clinical contexts

Chapter 4



Whole-genome sequencing and variant analysis of HPV16 infections

Pascal van der Weele[a,b], Chris JLM Meijer[b], and Audrey J King[a]

[a]National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Research, Diagnostics and Screening, Bilthoven, the Netherlands
[b]Vrije Universiteit–University Medical Center (VUmc), Department of Pathology, Amsterdam, the Netherlands

J Virol. 2017 Oct 1; 91(19)

Abstract

Human papillomavirus (HPV) is a strongly conserved DNA virus, of which high-risk types can cause cervical cancer in persistent infections. The most common type found in HPV-attributable cancer is HPV16, which can be subdivided into four lineages (A-D) with different carcinogenic properties. Studies have shown HPV16 sequence diversity in different geographical areas, but only limited information is available regarding HPV16 diversity within a population, especially at the whole genome level. We have analyzed HPV16 major variant diversity, conservation in persistent infections and performed a SNP comparison between persistent and clearing infections. Materials were obtained in the Netherlands from a cohort study with longitudinal follow-up for up to three years. Our analysis shows a remarkably large variant diversity in the population. Whole genome sequences were obtained for 57 persistent and 59 clearing HPV16 infections, resulting in 109 unique variants. Interestingly, persistent infections were completely conserved through time. One reinfection event was identified, where the initial and follow-up sample clustered differently. Non-A1/A2 variants seemed to clear preferentially ($p = 0.02$). Our analysis shows population-wide HPV16 sequence diversity is very large. In persisting infections the HPV16 sequence is fully conserved. Sequencing can identify HPV16 reinfections, although occurrence is rare. SNP comparison identified no strong acting effect of the viral genome affecting HPV16 infection clearance or persistence in up to three years of follow-up. These findings suggest progression of an early HPV16 infection could be host-related.

Importance

Human papillomavirus (HPV) type 16 is the predominant type found in cervical cancer. Progression of initial infection to cervical cancer has been linked to sequence properties; however, knowledge on variants circulating in European populations is limited, especially with longitudinal follow-up. By sequencing a number of infections with known follow-up for up to three years, we gain initial insights in the genetic diversity of HPV16 and effects of the viral genome on the persistence of infections. A SNP comparison between sequences obtained from clearing and persisting infections did not identify strong-acting DNA variations responsible for these infection outcomes. In addition, we have identified a HPV16 reinfection event, where sequencing of initial and follow-up samples showed different HPV16 variants. Based on conventional genotyping, this infection would incorrectly be considered a persistent HPV16 infection. In the context of vaccine efficacy and monitoring studies, such infections could potentially cause reduced reported efficacy or efficiency.

Introduction

Human papillomavirus infection (HPV) is one of the most common sexually transmitted infections worldwide (1) and the causative agent of cervical cancer (2). HPVs are highly conserved, double stranded DNA viruses that have co-diverged with human populations for millennia (3). Most infections regress naturally, but high-risk HPV (hrHPV) infections that are not cleared by the host can cause cervical intraepithelial neoplasia (CIN) and cancer. Of all hrHPV types, HPV16 is the most carcinogenic, causing over 60% of cervical cancers worldwide (4). Based on whole-genome sequence data, HPV16 can be subdivided into four lineages (A-D), with different carcinogenic potential and geographical heritage attributed to each lineage (5, 6). Lineages differ between 1.0–10% at the whole-genome nucleotide level and are further divided into sublineages if the nucleotide difference between two variants from the same lineage is 0.5–1.0% (7). Variants which match host ethnicity, in turn, have been associated with increased risk of persistence (8) and more recently with increased risk of CIN3+ (6, 8). Additionally, multiple HPV16 variant co-infections are common (5), although the role of minority variants to an infection is unknown. In this study we focus on identifying HPV16 diversity of major variants circulating in a Dutch population via Sanger whole-genome sequencing (WGS). Whole-genome sequence analysis could provide increased resolution over individual genes or genomic segments. Moreover, differences in phylogenetic clustering, SNP locations and participant ethnicity were compared for clearing and persisting infections in a longitudinal cohort study among young women in the Netherlands (16–29y).

Information about occurrence rates of type-specific (TS) reinfection events could be relevant in a vaccine context, where vaccine efficacy and efficiency are being reported based on conventional genotyping assays (9–11). A TS reinfection could be interpreted as a false positive persistent HPV16 infection and possibly lead to reduced reported vaccine efficacy if based solely on conventional genotyping results. Therefore, we sequenced persistent HPV16 infections with longitudinal follow-up to discriminate between true persistent HPV16 infections and type-specific (TS) HPV16 variant reinfection events, which have been shown to occur previously (12).

Results

In this study, 499 study participants (15.2%) were found to be HPV16 positive. Persistent infections were found in 176 participants (5.4%; Figure 1). Characteristics of the participant subsets included in this study are shown in Table 1 and subsets were found representative for respective total groups. Full genome sequences were initially obtained from 58 participants with clearing infections, resulting in 58 whole genome sequences. Complete genomes were obtained from 57 participants with persistent infections. At least one round of follow-up was sequenced for 40 persisting infections, with on average 70.3 weeks between the first and last available sample (min 40, max 148 weeks). An additional 17 persisting infections only had a single round sequenced, resulting in 108 genomes from persistent infections in total.

Phylogeny

Phylogenetic analysis of HPV16 genomes is shown in Figure 2. Out of 115 HPV16 infections, 109 unique genomic variants were identified, meaning most infections were caused by unique sequence variants. Many of the identified variants differ by less than ten nucleotides from each other (Figure 2). The majority of study participants were infected with HPV16 genome variants clustering near reference A strains. A1 was the best represented sublineage, with 79 variants. Sublineages A2 and A4 were represented by 19 and two variants resepectively, while sublineage A3 was not found within this dataset. A subset of study participants was found to be infected with HPV16 variants representing strains C (n = 6), D1 (n = 1), and D3 (n = 1). No infections were found to cluster with lineage B. Upon sequencing, one participant with a clearing infection showed a variant that did not cluster with any of the described reference strains. The closest reference strain was found to be A3, at 80 nucleotides difference. The high variant diversity largely lost when zooming in on individual genes or the upstream regulatory region (URR), implying variations occurring across the complete genome (data not shown). No clear difference could be identified between clearing and persisting infections based on phylogenetic comparison, however, participants infected with A4, C and D strains (n = 10) seemed to clear the infections preferentially as nine clearing infections were identified and only one persistent infection ($p = 0.02$).

Ethnicity of participants

Study participants were predominantly of European heritage, 86.0% (n = 49/57) and 79.7% (n = 47/59) among persisting and clearing infections respectively (Table 1). Only 14.0% of persisting infections were obtained from participants with non-European ethnicity (8.8% mixed, n = 5/57 and 5.2% Asian, n = 3/57). From clearing infections 19.3% were obtained from non-European study participants (11.8% mixed, n = 7/59 and 8.5% Asian, n = 5/59). No other ethnicities were reported in the sequenced subsets. The distribution of persistent and clearing infections was not affected by ethnicity for this study (Fisher's exact: $p = 0.62$). Due to low numbers of non-European participants in the dataset, statistical analysis matching variants with ethnicity is not possible.

Longitudinal sampling of persistent infections

Multiple whole-genome sequences were obtained from forty study participants with persistent HPV16 infections. Twenty-six infections were sequenced at two sampling moments, twelve at three moments and two at four moments (Figure 1). All but one sequence remained completely unchanged over time. One study participant, initially considered as persistently infected, was actually found to have a different HPV16 variant in the follow-up sample, implying a type-specific reinfection. The initial sample clustered near reference strain A1, while the follow-up sample clustered near D3 with 151 nucleotide differences between samples (Figure 2). Both samples from this study participant were resequenced and confirmed in an independent Illumina sequencing experiment (data not shown). Concordance between Sanger and Illumina consensus was > 99.8% for both samples.

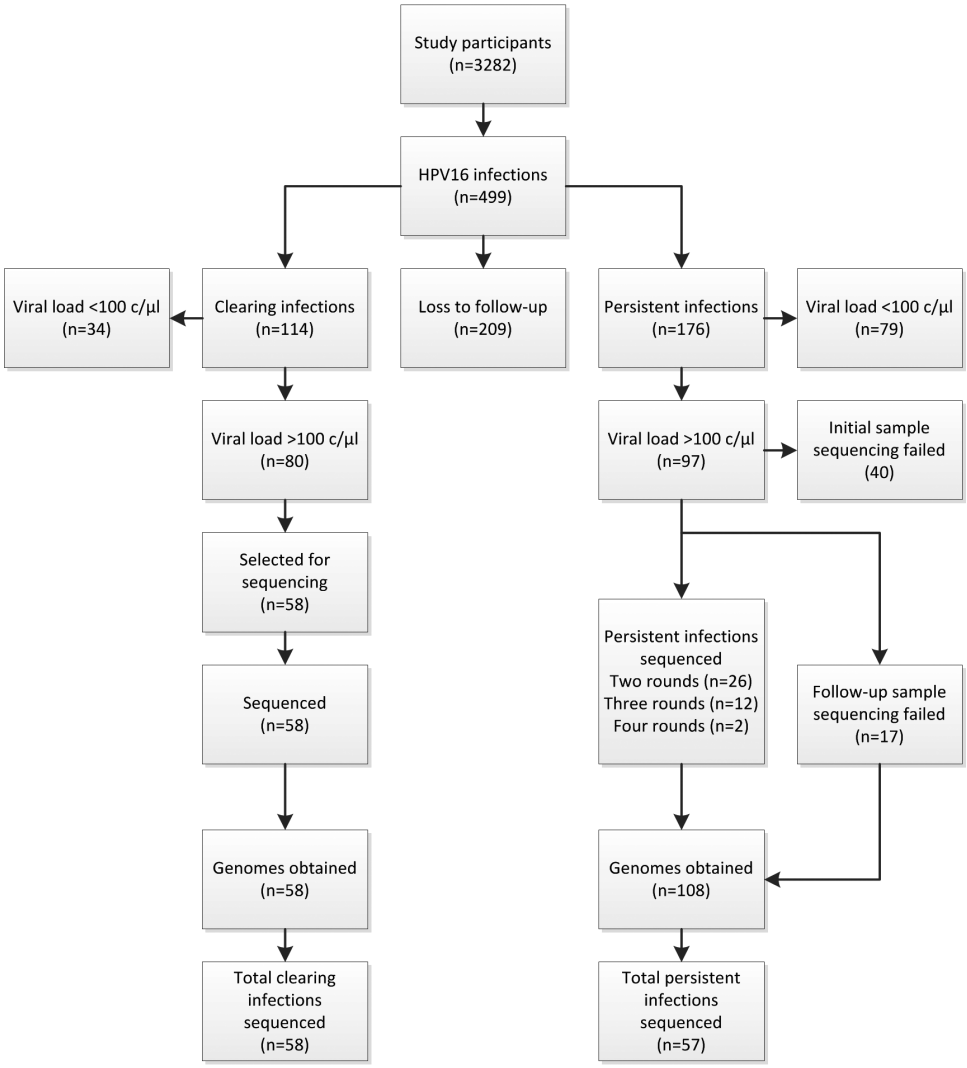


Figure 1: Schematic overview of selections made for persisting and clearing human papillomavirus 16 infections by duration (rounds), viral load criterion and sequencing results. Sequenced infections that persisted in three or four rounds had at least the initial and final samples sequenced. For clearing infections 58 genomes were obtained from 58 infections, for persisting infections 108 genomes were obtained from 57 infections.

Table 1: Characteristics of study and subsets of participants from who complete HPV16 genomes were obtained. Selected subsets were not found to be significantly different from the total group for each parameter ($p > 0.05$). Difference in age was assessed by Student's t-test, while differences in C. trachomatis status and ethnicity were assessed by Fisher's exact. The distributions of Dutch and non-Dutch variants in clearing and persistent infections were also compared and found to be non-significant (Fisher's exact: $p = 0.29$).

Characteristics of study subsets	Persistent infections sequenced	Clearing infections sequenced
	57/176	58/114
Age	Median years (95% CI)	
All persistent/clearing infections	25 (24-25)	23 (22-24)
Sequenced subset	24 (22-26)	22.5 (22-24)
***C. trachomatis* status**	*C. trachomatis* positive (n) / Total (n)	
All persistent/clearing infections	17/176	16/114
Sequenced subset	4/57	5/58
European ethnicity	European (n) / Total (n)	
All persistent/clearing infections	145/176	87/114
Sequenced subset	49/57	46/58
Mixed ethnicity	Mixed (n) / Total (n)	
All persistent/clearing infections	15/176	14/114
Sequenced subset	5/57	7/58
Asian ethnicity	Asian (n) / Total (n)	
All persistent/clearing infections	10/176	11/114
Sequenced subset	3/57	5/58
Other ethn-icities	Other (n) / Total (n)	
All persistent/clearing infections	6/176	2/114
Sequenced subset	0/57	0/58

HPV16 WGS based SNP analysis

In total 399 DNA SNPs, twelve insertions and seven deletions were identified when compared to reference strain K02178 across study participants (Table S2). Of all SNPs, 136 (34.1%) were found to lead to amino-acid (AA) changes (Table S2). None of the deletions or insertions were found in coding regions of the genome, except for one 63 nucleotide duplication in-frame in E1, which has been described previously (13).

With consideration of the aforementioned reinfection event, the final dataset consisted of 59 clearing and 56 persistent HPV16 infections. As non-A1/A2 variants were previously shown to clear preferentially, infections related to sublineages A1 and A2 were selected for SNP comparison. Non-A1/A2 strains were excluded from this analysis to prevent a bias in preferentially clearing SNPs from these variants. Participants infected with A1/A2 related strains (n = 105) were divided relatively evenly at 50 clearing and 55 persistent infections. SNP frequency and comparison between groups is shown in Figure 3. No significant coding differences were found among participants leading to preferential persisting or clearing of infections. One non-coding SNP was found significantly more often among study participants with clearing infections than with persisting infections (n = 18 vs n = 8; p = 0.048), at position 4185 in the E5-L2 intergenic region. Sliding window analysis shows similar patterns in nucleotide diversity between A1/A2 clearing and persistent infections (Figure 4).

Discussion

In this study we used whole genome sequencing to reveal a remarkably large population of unique HPV16 variants circulating naturally in young women in the Netherlands. The observed diversity between HPV16 variants is large enough to allow us to discriminate between true persistent HPV16 infections and variant reinfections in a longitudinal cohort study. Additionally, preferential clearing of variants belonging to (sub)lineages A4, C and D was observed. A large number of SNPs was identified in the present study when compared to reference strain K02718, however SNP diversity between clearing and persisting infections was non-significant for A1 and A2 related variants.

Phylogenetic analysis showed that most study participants were infected with variants clustering well with reference A lineages, as could be expected for a Dutch cohort (14). Among HPV16 genomes, one variant showed > 1.0% (80 nucleotides) sequence difference with the phylogenetically closest reference strain (A3). Based on definitions introduced by Burk *et al.* (7) this variant could be considered a new lineage.

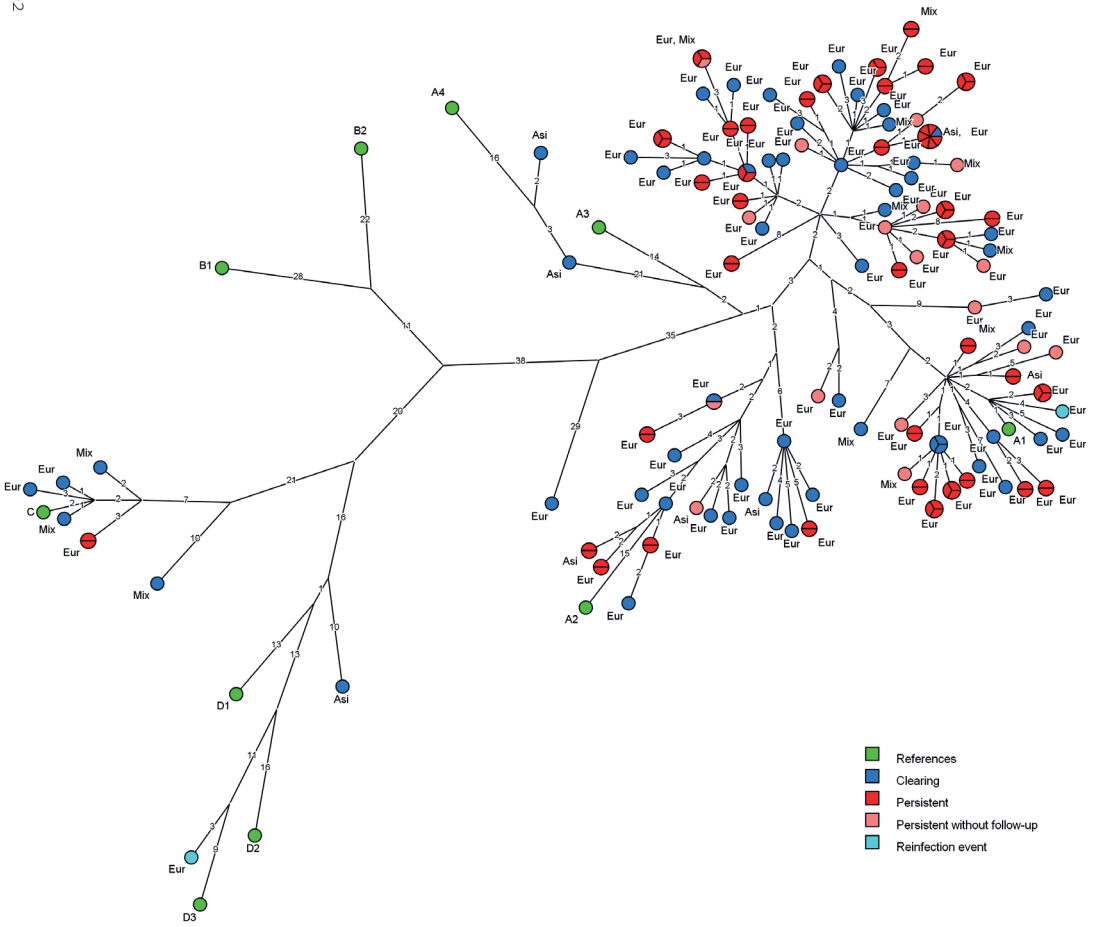


Figure 2: Maximum parsimony tree for human papillomavirus (HPV) 16 sequences. Each circle represents a specific variant, while parts within the circle show how often a variant occurs. In green are reference strains according to (7). In blue are clearing HPV16 infections. In red and pink are persistent infections with and without successfully sequenced follow-up samples. Variants to the right of the red line are considered part of A1/A2 sublineages, variants to the left of the red line are considered non-A1/A2 variants. Persistent infections had identical sequences at all sequenced timepoints. Light blue circles show a reinfection event that was initially considered a persisting infection based on conventional genotyping. Numbers on connecting lines indicate nucleotide difference between variants. Variants are coded C or P for clearing or persisting, followed by the number assigned to an infection and, for persistent infections, the round in which a variant was found. If no round information is specified, no follow-up data was obtained for these infections. Ethnicity of participants is indicated at variants using Eur for European, Asi for Asian and Mix for mixed heritage.

We found strong conservation of HPV16 variants within hosts with persisting HPV16 infections, indicating true persistency in the majority of infections. To our knowledge we are the first to show that no sequence variation occurs in the complete genome of persisting infections with an average follow-up of 70.3 weeks between initial and last available samples (min 40 weeks, max 148). Furthermore, we identified a HPV16 variant reinfection in one study participant, where the initial sample clustered differently from the follow-up sample. Indeed this implies that conventional genotyping could lead to a false-positive observation of persisting infections, although considering the occurrence in this dataset (1.8%, n = 1/57), conventional genotyping is very adequate at classifying persisting HPV16 infections. In the context of vaccine efficacy this could theoretically mean the difference between complete and partial reported protection, possibly justifying additional investigation in specific cases.

Despite the strong conservation of HPV16 sequences in persisting infections we find that nearly all participants with HPV16 infections are infected with unique variants, although variation between variants can be as little as one nucleotide. Single nucleotide differences between variants could be due to sequencing errors. However, in persisting infections we see that variants are completely conserved through time, even with very small differences between related variants. This precludes the possibility that we found artificial variants due to sequencing errors.

When combined, variants belonging to (sub)lineages A4, C and D seemed to clear preferentially compared to A1 and A2 variants ($p = 0.02$), but we could not perform additional statistics on combinations of variants and ethnicity, due to the very limited number of study participants with non-European ethnicities. For A1 and A2 related variants SNPs were assessed, as no phylogenetic distinction could be made between persistent and clearing infections. For these variants clearance and persistence were relatively equally distributed. Over the complete dataset a large number of SNPs was identified, but only one non-coding SNP was significantly different between groups. This SNP has no known biologically relevant function. In addition, no clear difference was found between nucleotide diversity for clearing or persistent infections. This might imply that kinetics for early HPV16 infection are mostly host-dependent, or at least less dependent on the viral genome. However, it could also mean that for the present study, the dataset is too small to identify differences within phylogenetically related lineages. Our SNP comparison implies that any variations between infections are mere effects of chance under minimal evolutionary pressure for this dataset. The source of this diversity might be ancient considering papillomavirus evolution (3).

This study has a number of limitations. Firstly, our definitions of persistent and clearing infections are not ideal. Due to The lack of information about the HPV status at baseline clearing infections as defined in this study might potentially be a mixture of true incident and clearing persistent infections. In addition, we do not know the time persistent infections could have been present before detected in the present study, or how long they would remain after the end of study. Therefore we can not exclude that persistent infections we identify in this study, ultimately clear. This might explain why no strong-acting differences at SNP level between persistent and clearing infections were identified. Unfortunately, because this study was designed as a Chla-

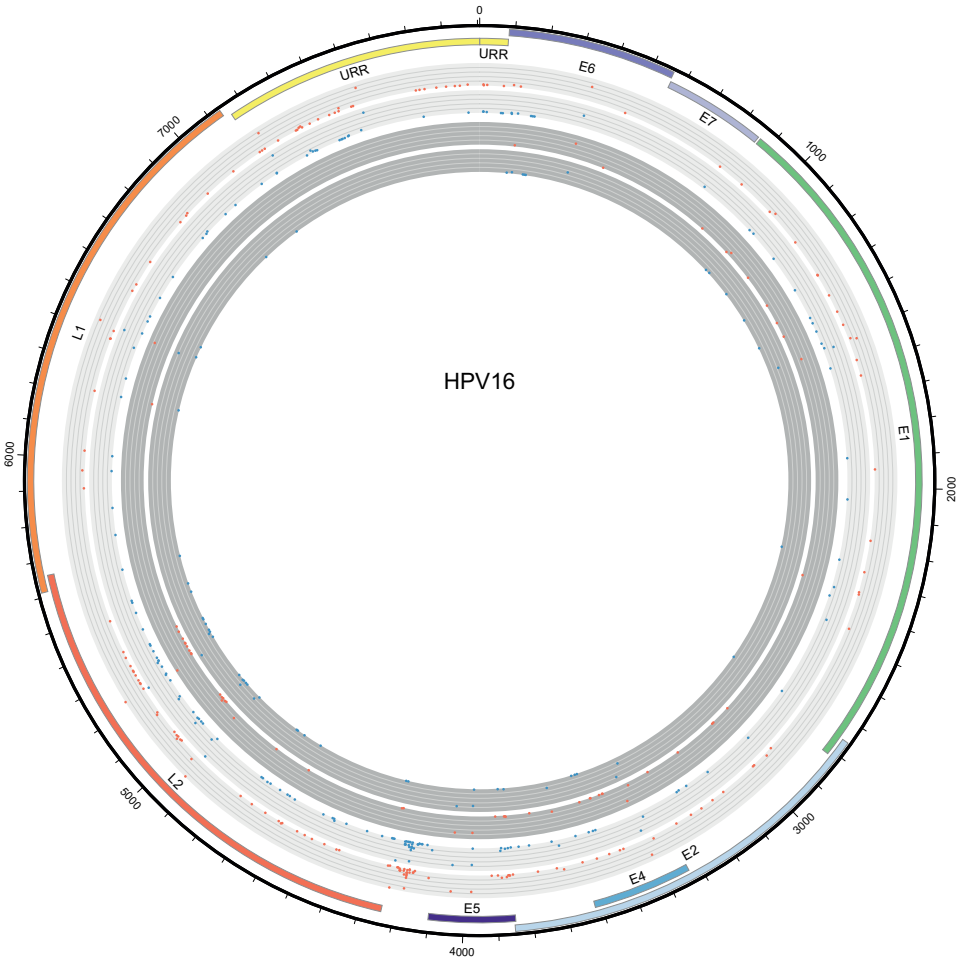


Figure 3: Circular representation of single-nucleotide polymorphism comparison on the human papillomavirus (HPV) 16 genome. Light grey circles show DNA variations, while dark grey circles show amino-acid changes. Changes found among clearing (n = 50) and persisting (n = 55) variants are shown in blue and red respectively. Height of SNPs on respective circles denotes relative incidence of variations in the dataset compared to reference strain K02718.

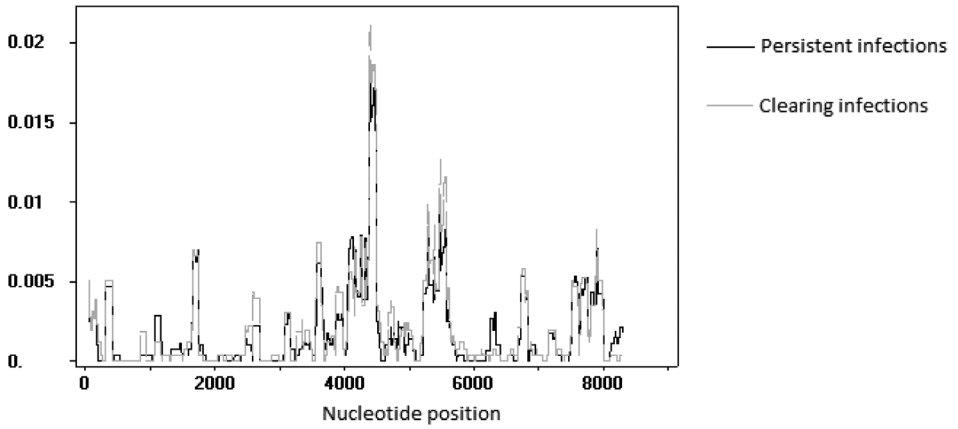


Figure 4: Plot of nucleotide diversity (pi) across A1/A2 HPV16 genome based on sliding window analysis. Data from persistent (n = 55) and clearing (n = 50) infections is represented in black and grey respectively.

mydia screening study no cytologic or histologic follow-up is available. Therefore we defined clearing infection as the apparent lack of HPV16 detection in follow-up samples. Our data does not discriminate between the various ways HPV infections could be cleared.

Although our dataset shows great diversity among the population, it is relatively small for assessment of differences between clearing and persistent infections, especially if these effects are not very pronounced. Due to the study size, only very strong differences driving persistence or clearance of infection can be identified. In our dataset we do not identify a clear link between specific DNA variations and infection outcome, but it could be possible that these effects are more nuanced for individual SNPs. Such effects would be beyond this dataset and further research on larger datasets is required to identify if these effects could drive persistence of HPV16 infection.

Sanger WGS was accomplished by amplification of two 4.5 kb fragments covering the full genome. A viral load of at least 100 c/µl was empirically determined as the minimum concentration of HPV where full-length sequencing results could be expected. This could lead to a possible enrichment of variants resulting in high viral load infections.

Although Sanger sequencing is still the golden standard, it lacks the resolution generated by next-generation sequencing (NGS) techniques. Only the major variant driving the infection can be reliably determined, any co-infections present in the samples can not be reliably identified (15). On the other hand, major variants have been implicated to cause persistent infections, while minor variants appear more transient in nature (16). If a second variant is present in a sample at a concentration comparable to that of the major variant, Sanger sequencing could result in double peaks at certain nucleotide positions. In these instances the software base-calling algorithm supplemented with manual verification were used to reach a definitive consensus. In this dataset double peaks were rare and generally limited to a single read, where other reads at the same position did result in single clear curves.

Additionally, reinfections with the same variant could not be identified using our method. This could occur when the partner of the person sampled in this study carries a persisting infection, causing a 'ping-pong' effect in which repeated exposure to the same variant occurs. It remains to be seen if even the increased resolution NGS provides could be of discriminative value for such cases. Further research utilizing NGS will be required to assess the role of minority variants in persistent infections.

For this study no long-term follow-up is available. Study participants with infections that were identified as persistent, might actually have slow clearing infections. This could lead to a number of clearing infections in the persistent group, possibly impeding identification of SNPs truly associated with persistence of infection. Infections that were positive for two rounds, followed by a negative sample were considered, but only two were identified in all of the sequenced data, preventing any analysis.

In summary, we have applied whole genome sequencing to show that HPV16 variants in the Netherlands are highly diverse between study participants, but conserved through time in persisting infections. SNP analysis shows a large number of variable sites, but no clear differences between clearing or persisting infections. This might imply that infection persistence at an early stage is weakly mediated by the virus and possibly more host related. Reinfection events can occur, albeit very rarely in the population. In the context of vaccine efficiency studies our results provide useful information about the behavior of HPV16 infections through time and may be of use in monitoring vaccine efficiency.

Materials and Methods

CSI study design

Vaginal self-swabs were collected from participants of the *Chlamydia trachomatis* Screening and Implementation Program (CSI). Study recruitment and methods have been described previously (17, 18). 3282 participants that gave additional consent for STI testing other than *C. trachomatis* and answered a questionnaire were included in this study (19). Ethnicity was based on country of birth of study participants and their parents and assigned according to Woestenberg *et al.* (20). Ethnicity was divided in European, Asian, 'mixed' (participants from the Caribbean and surrounding areas), and 'other' ethnicities (combining all other nationalities reported for HPV16 positive study participants). Study participants supplied samples in up to four rounds each, with median fifty weeks between rounds (95%CI: 49–50 weeks, min: 5, max: 101 weeks). This study was approved by the Medical Ethical Committee of VUmc Amsterdam (2007/239).

HPV DNA detection, genotyping and quantification

Sample DNA isolation and HPV DNA genotyping have been described previously (19). In short, total DNA was extracted from 200 µl of vaginal swab using the MagNA Pure 96 platform (Total Nucleic Acid Isolation Kit, Roche Diagnostics) according to the manufacturer's protocol and eluted in 100 µl. Genotyping occurred using the SPF10-DEIA-LiPA$_{25}$ platform (DDL Diagnostics) (21, 22). Samples positive for HPV16 were quantitated previously (23).

Sample selection criteria

An arbitrary PCR threshold was empirically defined at a viral load of 100 copies / µl (c/µl). Samples with HPV DNA concentrations below this value were considered likely to fail in the PCR step and were therefore not analyzed. Persistent HPV16 infections with the first and last samples above the viral load threshold were selected for WGS analysis. For infections persisting three or four rounds, at least the initial and last samples were sequenced. Persistent infections were defined as TS HPV16 positive in two or more subsequent rounds with at least forty weeks between samples. Infections with follow-up less than forty weeks were excluded (n = 1). In addition, to reach equal amounts of persistent and clearing infections, HPV16 positive samples were randomly selected from all participants with clearing HPV16 infections that met viral load criteria (Figure 1). Clearing infections were defined as HPV16 positive in the round of sequencing, followed by a HPV16 negative test.

Long-template PCR and sequencing

DNA eluates were subjected to long-template PCR to amplify the complete HPV16 genome. Two overlapping fragments, encompassing the complete genome, were generated using primer combinations F1832 + R6382 and F6201 + R2915 (24) (Table S1). PCR was performed using Takara PrimeSTAR GXL according to the manufacturer's protocol. Cycling conditions consisted of an initial incubation at 98°C for eight minutes followed by 38 cycles of 98°C denaturation for 15 seconds, 55°C annealing for 30 seconds and 68°C elongation for five minutes, followed by a final elongation step at 68°C for 15 minutes.

PCR product amplification was verified on the Lonza FlashGel (Lonza) system. If both fragments amplified successfully, samples were treated with ExoSAP-IT PCR product clean-up (Affymetrix) according to the manufacturer's protocol. If amplification failed for the initial sample, the follow-up sample was excluded from further analyses. If amplification succeeded for the initial sample, but failed for the follow-up sample, these infections were sequenced without follow-up. Purified PCR products were subjected to Sanger sequencing using 45 unique primers for HPV16, covering the complete genome in both forward and reverse direction (24–28) (Table S1).

Whole-genome sequencing and phylogenetic analyses

Sanger WGS data obtained was analyzed using CLC Genomics Workbench 9.5.3 (CLC Bio, Qiagen). For each sample, reads were assembled against reference strain K02718 (29). Assembled genomes with coverage < 1 at any nucleotide position were omitted from analysis. A consensus was generated based on assembled reads. Obtained sequences were verified manually in the assembly to compensate for possible base-calling errors by the software algorithm. Upon finalization of consensus, sequences were exported to BioNumerics 7.2.5 (AppliedMaths) as GenBank (.gbk) files for phylogenetic analysis.

Reference strains used in phylogenetic analyses were selected based on (7). Represented HPV16 lineages and sublineages were A1–4, B1–2, C and D1–3.

If reinfection events were found within the longitudinal analysis of persistent infections, the initial sample of the infection was at that point regrouped under clearing infections for downstream analysis. The follow-up sample was treated accordingly depending on available further follow-up.

SNP and statistical analysis

For all samples single nucleotide polymorphisms (SNPs), amino-acid (AA) changes, insertions and deletions were analyzed with respect to reference strain K02718, using ProSeq 3.5. Coding regions for HPV16 genes were used according to Papilloma Virus Episteme (30) (PaVE, pave.niaid.nih.gov). SNP comparison was visualized using Circos (www.circos.ca). Nucleotide diversity (pi) between persistent and clearing infections was assessed using DNAsp with a sliding window size of 100 and a step size of 1. Differences in individual SNP occurrence between clearing and persistent infections were assessed using Fisher's exact test. To compensate for possible sequence or interpretation errors, only SNPs occurring > 1 times in the complete dataset were considered for further investigation.

Accession numbers

Sequences obtained in this study were submitted to Genbank (accession numbers: KY549156-KY549321).

References

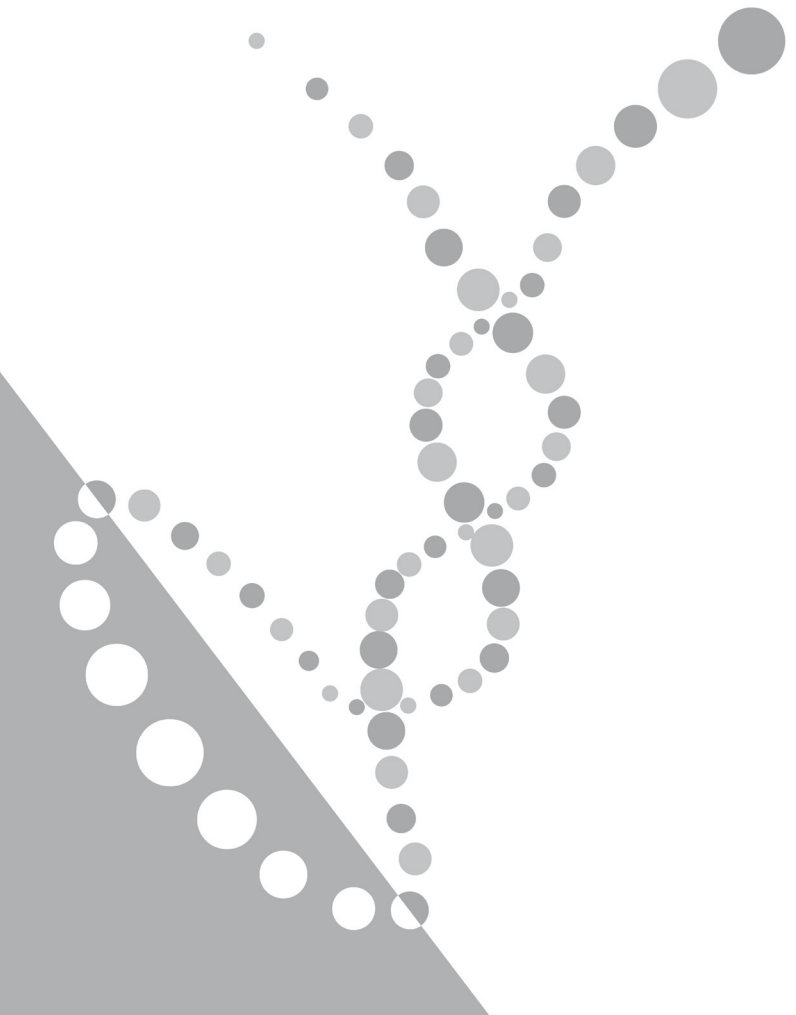
1. Baseman JG, Koutsky LA. 2005. The epidemiology of human papillomavirus infections. J Clin Virol 32 Suppl 1:S16–24.
2. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Munoz N. 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 189:12–9.
3. Van Doorslaer K. 2013. Evolution of the papillomaviridae. Virology 445:11–20.
4. de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, Tous S, Felix A, Bravo LE, Shin HR, Vallejos CS, de Ruiz PA, Lima MA, Guimera N, Clavero O, Alejo M, Llombart-Bosch A, Cheng-Yang C, Tatti SA, Kasamatsu E, Iljazovic E, Odida M, Prado R, Seoud M, Grce M, Usubutun A, Jain A, Suarez GA, Lombardi LE, Banjo A, Menendez C, Domingo EJ, Velasco J, Nessa A, Chichareon SC, Qiao YL, Lerma E, Garland SM, Sasagawa T, Ferrera A, Hammouda D, Mariani L, Pelayo A, Steiner I, Oliva E, Meijer CJ, Al-Jassar WF, Cruz E, Wright TC, Puras A, et al. 2010. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol 11:1048–56.
5. Cullen M, Boland JF, Schiffman M, Zhang X, Wentzensen N, Yang Q, Chen Z, Yu K, Mitchell J, Roberson D, Bass S, Burdette L, Machado M, Ravichandran S, Luke B, Machiela MJ, Andersen M, Osentoski M, Laptewicz M, Wacholder S, Feldman A, Raine-Bennett T, Lorey T, Castle PE, Yeager M, Burk RD, Mirabello L. 2015. Deep sequencing of HPV16 genomes: A new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection. Papillomavirus Research doi:http://dx.doi.org/10.1016/j.pvr.2015.05.004.
6. Mirabello L, Yeager M, Cullen M, Boland JF, Chen Z, Wentzensen N, Zhang X, Yu K, Yang Q, Mitchell J, Roberson D, Bass S, Xiao Y, Burdett L, Raine-Bennett T, Lorey T, Castle PE, Burk RD, Schiffman M. 2016. HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women. J Natl Cancer Inst 108.
7. Burk RD, Harari A, Chen Z. 2013. Human papillomavirus genome variants. Virology 445:232–43.
8. Xi LF, Kiviat NB, Hildesheim A, Galloway DA, Wheeler CM, Ho J, Koutsky LA. 2006. Human papillomavirus type 16 and 18 variants: race-related distribution and persistence. J Natl Cancer Inst 98:1045–52.
9. Apter D, Wheeler CM, Paavonen J, Castellsague X, Garland SM, Skinner SR, Naud P, Salmeron J, Chow SN, Kitchener HC, Teixeira JC, Jaisamrarn U, Limson G, Szarewski A, Romanowski B, Aoki FY, Schwarz TF, Poppe WA, Bosch FX, Mindel A, de Sutter P, Hardt K, Zahaf T, Descamps D, Struyf F, Lehtinen M, Dubin G, Group HPS. 2015. Efficacy of human papillomavirus 16 and 18 (HPV-16/18) AS04-adjuvanted vaccine against cervical infection and precancer in young women: final event-driven analysis of the randomized, double-blind PATRICIA trial. Clin Vaccine Immunol 22:361–73.
10. Joura EA, Giuliano AR, Iversen OE, Bouchard C, Mao C, Mehlsen J, Moreira ED, Jr., Ngan Y, Petersen LK, Lazcano-Ponce E, Pitisuttithum P, Restrepo JA, Stuart G, Woelber L, Yang YC, Cuzick J, Garland SM, Huh W, Kjaer SK, Bautista OM, Chan IS, Chen J, Gesser R, Moeller E, Ritter M, Vuocolo S, Luxembourg A, Broad Spectrum HPVVS. 2015. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. N Engl J Med 372:711–23.
11. Struyf F, Colau B, Wheeler CM, Naud P, Garland S, Quint W, Chow SN, Salmeron J, Lehtinen M, Del Rosario-Raymundo MR, Paavonen J, Teixeira JC, Germar MJ, Peters K, Skinner SR, Limson G, Castellsague X, Poppe WA, Ramjattan B, Klein TD, Schwarz TF, Chatterjee A, Tjalma WA, Diaz-Mitoma F, Lewis DJ, Harper

DM, Molijn A, van Doorn LJ, David MP, Dubin G, Group HPS. 2015. Post hoc analysis of the PATRICIA randomized trial of the efficacy of human papillomavirus type 16 (HPV-16)/HPV-18 AS04-adjuvanted vaccine against incident and persistent infection with nonvaccine oncogenic HPV types using an alternative multiplex type-specific PCR assay for HPV DNA. Clin Vaccine Immunol 22:235–44.

12. Xi LF, Koutsky LA, Castle PE, Edelstein ZR, Hulbert A, Schiffman M, Kiviat NB. 2010. Human papillomavirus type 16 variants in paired enrollment and follow-up cervical samples: implications for a proper understanding of type-specific persistent infections. J Infect Dis 202:1667–70.
13. Sabol I, Matovina M, Si-Mohamed A, Grce M. 2012. Characterization and whole genome analysis of human papillomavirus type 16 e1–1374^63nt variants. PLoS One 7:e41045.
14. Yamada T, Manos MM, Peto J, Greer CE, Munoz N, Bosch FX, Wheeler CM. 1997. Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. J Virol 71:2463–72.
15. Shen-Gunther J, Wang Y, Lai Z, Poage GM, Perez L, Huang TH. 2017. Deep sequencing of HPV E6/E7 genes reveals loss of genotypic diversity and gain of clonal dominance in high-grade intraepithelial lesions of the cervix. BMC Genomics 18:231.
16. Quint W, Jenkins D, Molijn A, Struijk L, van de Sandt M, Doorbar J, Mols J, Van Hoof C, Hardt K, Struyf F, Colau B. 2012. One virus, one lesion—individual components of CIN lesions contain a specific HPV type. J Pathol 227:62–71.
17. van den Broek IV, Hoebe CJ, van Bergen JE, Brouwers EE, de Feijter EM, Fennema JS, Gotz HM, Koekenbier RH, van Ravesteijn SM, de Coul EL. 2010. Evaluation design of a systematic, selective, internet-based, Chlamydia screening implementation in the Netherlands, 2008–2010: implications of first results for the analysis. BMC Infect Dis 10:89.
18. van den Broek IV, van Bergen JE, Brouwers EE, Fennema JS, Gotz HM, Hoebe CJ, Koekenbier RH, Kretzschmar M, Over EA, Schmid BV, Pars LL, van Ravesteijn SM, van der Sande MA, de Wit GA, Low N, Op de Coul EL. 2012. Effectiveness of yearly, register based screening for chlamydia in the Netherlands: controlled trial with randomised stepped wedge implementation. BMJ 345:e4316.
19. Mollers M, Boot Hein J, Vriend Henrike J, King Audrey J, van den Broek Ingrid VF, van Bergen Jan EA, Brink Antoinette AT, Wolffs Petra FG, Hoebe Christian JP, Meijer Chris JL, van der Sande Marianne AB, de Melker Hester E. 2013. Prevalence, incidence and persistence of genital HPV infections in a large cohort of sexually active young women in the Netherlands. Vaccine 31:394–401.
20. Woestenberg PJ, van Oeffelen AA, Stirbu-Wagner I, van Benthem BH, van Bergen JE, van den Broek IV. 2015. Comparison of STI-related consultations among ethnic groups in the Netherlands: an epidemiologic study using electronic records from general practices. BMC Fam Pract 16:70.
21. Kleter B, van Doorn LJ, Schrauwen L, Molijn A, Sastrowijoto S, ter Schegget J, Lindeman J, ter Harmsel B, Burger M, Quint W. 1999. Development and clinical evaluation of a highly sensitive PCR-reverse hybridization line probe assay for detection and identification of anogenital human papillomavirus. J Clin Microbiol 37:2508–17.
22. Kleter B, van Doorn LJ, ter Schegget J, Schrauwen L, van Krimpen K, Burger M, ter Harmsel B, Quint W. 1998. Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. Am J Pathol 153:1731–9.
23. van der Weele P, van Logchem E, Wolffs P, van den Broek I, Feltkamp M, de Melker H, Meijer CJ, Boot H, King AJ. 2016. Correlation between viral load, multiplicity of infection, and persistence of HPV16 and HPV18 infection in a Dutch cohort of young women. J Clin Virol 83:6–11.

24. Lurchachaiwong W, Junyangdikul P, Payungporn S, Chansaenroj J, Sampathanukul P, Tresukosol D, Termrungruanglert W, Theamboonlers A, Poovorawan Y. 2009. Entire genome characterization of human papillomavirus type 16 from infected Thai women with different cytological findings. Virus Genes 39:30–8.
25. Seaman WT, Andrews E, Couch M, Kojic EM, Cu-Uvin S, Palefsky J, Deal AM, Webster-Cyriaque J. 2010. Detection and quantitation of HPV in genital and oral tissues and fluids by real time PCR. Virol J 7:194.
26. Sun M, Gao L, Liu Y, Zhao Y, Wang X, Pan Y, Ning T, Cai H, Yang H, Zhai W, Ke Y. 2012. Whole genome sequencing and evolutionary analysis of human papillomavirus type 16 in central China. PLoS One 7:e36577.
27. Cornut G, Gagnon S, Hankins C, Money D, Pourreaux K, Franco EL, Coutlee F, Canadian Women's HIVSG. 2010. Polymorphism of the capsid L1 gene of human papillomavirus types 31, 33, and 35. J Med Virol 82:1168–78.
28. Yue Y, Yang H, Wu K, Yang L, Chen J, Huang X, Pan Y, Ruan Y, Zhao Y, Shi X, Sun Q, Li Q. 2013. Genetic variability in L1 and L2 genes of HPV-16 and HPV-58 in Southwest China. PLoS One 8:e55204.
29. Seedorf K, Krammer G, Durst M, Suhai S, Rowekamp WG. 1985. Human papillomavirus type 16 DNA sequence. Virology 145:181–5.
30. Van Doorslaer K, Tan Q, Xirasagar S, Bandaru S, Gopalan V, Mohamoud Y, Huyen Y, McBride AA. 2013. The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. Nucleic Acids Res 41:D571–8.

Chapter 5



High Whole-Genome Sequence Diversity of Human Papillomavirus Type 18 Isolates

Pascal van der Weele [1,2], Chris J.L.M. Meijer [2] and Audrey J. King [1]

[1]National Institute for Public Health and the Environment (RIVM),
Centre for Infectious Disease Control, 3721 MA Bilthoven, The Netherlands
[2]Department of Pathology, Vrije Universiteit—University Medical Center (VUmc),
1081 HV Amsterdam, The Netherlands

Viruses. 2018 Feb; 10(2): 68

Abstract

Background The most commonly found human papillomavirus (HPV) types in cervical cancer are HPV16 and HPV18. Genome variants of these types have been associated with differential carcinogenic potential. To date, only a handful of studies have described HPV18 whole genome sequencing results. Here we describe HPV18 variant diversity and conservation of persistent infections in a longitudinal retrospective cohort study.

Methods Cervical self-samples were obtained annually over four years and genotyped on the SPF10-DEIA-LiPA$_{25}$ platform. Clearing and persistent HPV18 positive infections were selected, amplified in two overlapping fragments, and sequenced using 32 sequence primers.

Results Complete viral genomes were obtained from 25 participants with persistent and 26 participants with clearing HPV18 infections, resulting in 52 unique HPV18 genomes. Sublineage A3 was predominant in this population. The consensus viral genome was completely conserved over time in persistent infections, with one exception, where different HPV18 variants were identified in follow-up samples.

Conclusions This study identified a diverse set of HPV18 variants. In persistent infections, the consensus viral genome is conserved. The identification of only one HPV18 infection with different major variants in follow-up implies that this is a potentially rare event. This dataset adds 52 HPV18 genome variants to Genbank, more than doubling the currently available HPV18 information resource, and all but one variant are unique additions.

Introduction

Persistent high-risk (hr) human papillomavirus (HPV) infection is the prime cause for cervical cancer [1]. Currently, 13 HPV types are considered to be high-risk based on criteria defined by the International Agency for Research on Cancer [2], with different carcinogenic risk for each hrHPV type. HPV16 and HPV18 cause approximately 70% of all cervical cancers worldwide [3], making them the primary target for research and vaccination alike.

HPVs are a strongly conserved dsDNA family of viruses. Individual HPV types are considered to show >90% sequence homology. HPV sequences with <10% nucleotide difference are considered the same HPV type, but can be subdivided into lineages (1–10% sequence difference) and sublineages (0.5–1.0% sequence difference). All sequences belonging to the same HPV type are considered variants [4]. Differential risks for cervical intraepithelial neoplasia (CIN) and cervical cancer have been analyzed for all hrHPV types [5–8]. A correlation between lineages and carcinogenicity has been described repeatedly for HPV16 [5–7, 9] and conservation of HPV16 E7 has been associated with carcinogenesis [10]. For HPV18, three lineages and eight sublineages have been described (A1–5, B1–3 and C) [4]. However, unlike HPV16, some ambiguity exists in the current research on the carcinogenic properties of variant lineages for HPV18 [5, 6, 8]. For other hrHPV types, the literature on lineages and carcinogenicity is limited. Only one large study has been conducted describing some type-specific (TS) correlations [11]. The studies described have succeeded in establishing the importance of HPV variant identification in the context of persisting infections, CIN lesions and finally cervical cancer. Research on differences between persistent infections and clearing infections at the viral genome level is limited [11, 12], especially for HPV18. High variant diversity has been described for HPV16 [9, 13], but literature describing HPV18 diversity based on whole-genome sequencing (WGS) within populations is scarce [14–16]. Here we present the HPV18 variant diversity in a longitudinal study population in the Netherlands using a WGS approach. Additionally, we assessed the prevalence of HPV18 infections where a shift occurs in major variants after follow-up, which could be considered reinfection events and which would incorrectly be considered persistent infections based on conventional genotyping. These findings could be relevant in a context where vaccine efficacy and efficiency is determined [17, 18].

Materials and Methods

Study Design

Samples were obtained from the *Chlamydia trachomatis* Screening and Implementation (CSI) study. This study was approved by the Medical Ethical Committee of VUmc Amsterdam (2007/239, METC VUmc, 2007) and registered in the Dutch Trial Registry (NTR3071, 16 Sept 2011). Recruitment criteria and methods have been previously described [19, 20], along with additional consent required for HPV testing [21]. 3282 study participants consented to additional HPV testing and supplied samples in up to four round, with median fifty weeks between rounds (95%CI: 49–50 weeks, min: 5, max: 101 weeks). Ethnicity was based on country of birth and assigned according to Woestenberg *et al*. [22]. Participants were divided in European, Asian, African and "mixed"

(Caribbean and Middle-American) ethnicities. Clearing infections were defined as type-specific HPV positive at the initial time point, followed by a type-specific HPV negative measurement. Persistent infections were type-specific HPV positive in at least two subsequent sampling rounds. Persistent infections were assessed for different major variants by sequence and phylogenetic comparison of initial and follow-up samples. If an infection was found to have different major variants at two subsequent time points, it was omitted from statistical (but not phylogenetic) analysis, since Sanger cannot reliably distinguish between persistent and clearing in these cases, due to limited sequence resolution.

DNA Purification and HPV Detection, Genotyping and Quantification

Total DNA was isolated from 200 µL of cervical self-swabs using the MagnaPure96 platform (Total Nucleic Acid Isolation Kit, Roche Diagnostics, Basel, Switzerland) according to the manufacturer's protocol. Isolates were eluted in 100 µL and subsequently genotyped via the SPF10-DEIA-LiPA$_{25}$ platform (DDL Diagnostics, Rijswijk, The Netherlands) [23, 24]. HPV18 positive samples were quantified previously [25].

Sample Selection, Long-Template PCR and Sequencing

HPV18 positive samples with viral load > 100 copies/µL (c/µL) were identified. Samples below this empirically defined threshold were likely to fail during amplification and therefore excluded. Persistent infections according to genotyping results were selected if the initial and last available follow-up samples were above the viral load criterium. In addition, 35 clearing HPV18 infections with VL above the arbitrary threshold were selected at random. Selected samples were subjected to an initial PCR amplifying the complete HPV18 genome in two overlapping fragments (F2458 + R6668 and F6538 + R3393 [15]). Amplification was performed using PrimeSTAR GXL DNA polymerase (Takara Bio, Kusatsu, Shiga, Japan). Cycling conditions consisted of an initial incubation at 98 °C for eight minutes followed by 38 cycles of alternating 98 °C for 15 s, 55 °C for 30 s and 68 °C for five minutes. A final elongation of 15 min at 68 °C was included. Amplification was verified on the FlashGel system (Lonza, Basel, Switzerland). Successfully generated amplicons were treated with ExoSAP-IT PCR product clean-up (Affymetrix, Santa Clara, CA, USA) according to manufacturer's protocol. If the initial sample of a persistent infection failed to amplify, the follow-up sample was omitted from further analysis. If the initial sample succeeded, but the follow-up sample failed to amplify, these infections were included in phylogenetic analysis, but excluded from statistical and single-nucleotide polymorphism (SNP) analysis as true persistency could not be confirmed. The complete HPV18 genome was Sanger sequenced using 31 unique sequencing primers [15, 26, 27] (Table S1).

Phylogenetic and SNP Analysis

Sequence data were assembled against HPV18 reference strain AY262282, updated from X05015 [28] using CLC Genomics Workbench 9.5.3 (CLC Bio, Aarhus, Denmark; Qiagen, Hilden, Germany). Coverage > 1 across the complete genome was required, otherwise samples were omitted. After manual verification to compensate for possible base-calling errors, consensus sequences were generated and exported to BioNumerics 7.2.5 (AppliedMaths, Sint-Martens-Latem, Belgium) for alignment and maximum parsimony phylogenetic analysis.

In addition, Multiple Sequence Comparison by Log-Expectation (MUSCLE) 3.8.31 was used to generate a multiple sequence alignment using standard settings. A maximum likelihood phylogenetic tree was generated using IQtree 1.6.1 with 1000 ultrafast bootstrap replicates (–bb 1000). An optimal substitution model was identified (–m MF) and resulted in the Kimura 3-parameter substitution model with empirical base frequencies and allowing for a proportion of invariable sites (–m K3Pu + F + I). The resulting Newick tree file was visualized using Figtree 1.4.3. Phylogenetic analyses were performed using reference strains for lineages (A, B and C) and sublineages (A1–5, B1–3) as described by [4].

SNPs were assessed using a consensus sequence generated from all sequences obtained in this study. Am SNP was considered orphan if it was found in only one participant at a single time point. Orphan SNPs were excluded from SNP analysis and comparison.

Statistical Analysis

General study characteristics were assessed by unpaired t or Fisher's exact testing. Differences in SNP prevalence in clearing or persisting infections were analyzed by Fisher's exact testing. Two-tailed p values < 0.05 were considered to be statistically significant. Nucleotide diversity (pi) for HPV18 clearing and persisting infections was calculated using a MUSCLE multiple alignment in DNAsp with a sliding window size of 100 nucleotides and a step size of one.

Accession Numbers

All HPV18 sequences were deposited to the Genbank database under accession numbers: MF288652-MF288727.

Results

Out of 3282 participants initially included in the CSI study, 259 were HPV18 positive (7.9%, Figure 1). Clearing and persisting infections were found in 79 (2.4%) and 66 (2.0%) study participants respectively. No follow-up information was available from 114 participants (3.5%).

Complete HPV18 genomes were obtained from 26 participants with clearing infections and 25 participants with persistent infections (Figure 1). From the 25 persistent infections, 21 were completely sequenced with at least one round of follow-up. Persistent infections had median 52.5 weeks (min 8, max 102 weeks) between sampling moments. Four persistent infections had the initial sample sequenced, but sequencing analysis failed in the follow-up. Across clearing and persisting infections combined, a total of 76 complete HPV18 sequences were obtained. General characteristics of the study subsets are presented in Table 1.

Phylogeny

A maximum parsimony phylogenetic tree of the HPV18 data was generated using BioNumerics and is shown in Figure 2. As could be expected for a Dutch cohort, most infections clustered near lineage A reference strains. A majority of the participants was infected with HPV18 sublineage A3 (64.7%, n = 33/51). Two participants were infected with variants clustering near the maximum allowed nucleotide difference for sublineages (0.47% and 0.49% from A3 respectively, while variants within a sublineage differ ≤0.5% at the whole genome level, [4]), implying relatively large diversity can occur within a sublineage (marked with * in Figure 2 and Figure S1). Sublineage A1 was isolated and sequenced from 21.6% (n = 11/51) of study participants, while sublineage A2 was found only once. Lineage B HPV18 infections were also detected in this study. Five infections clustered close to sublineage B1 and one to sublineage B2. Lineage C and sublineages A4 and B3 were not found in this study. In addition, a maximum likelihood analysis was performed using MUSCLE and IQtree, which showed highly similar clustering across all (sub-) lineages (Figure S1). Both trees showed an apparent overrepresentation of clearing infections in the A3 clade (18, compared to nine persistent, excluding two sequences marked with * in Figure 2 and Figure S1) and an apparent underrepresentation of persisting infections in the A1 clade (six, compared to three clearing), however, these differences were not significant (Fisher's exact, p = 0.13), even upon inclusion of persistent infections for which no follow-up could be sequenced (p = 0.15).

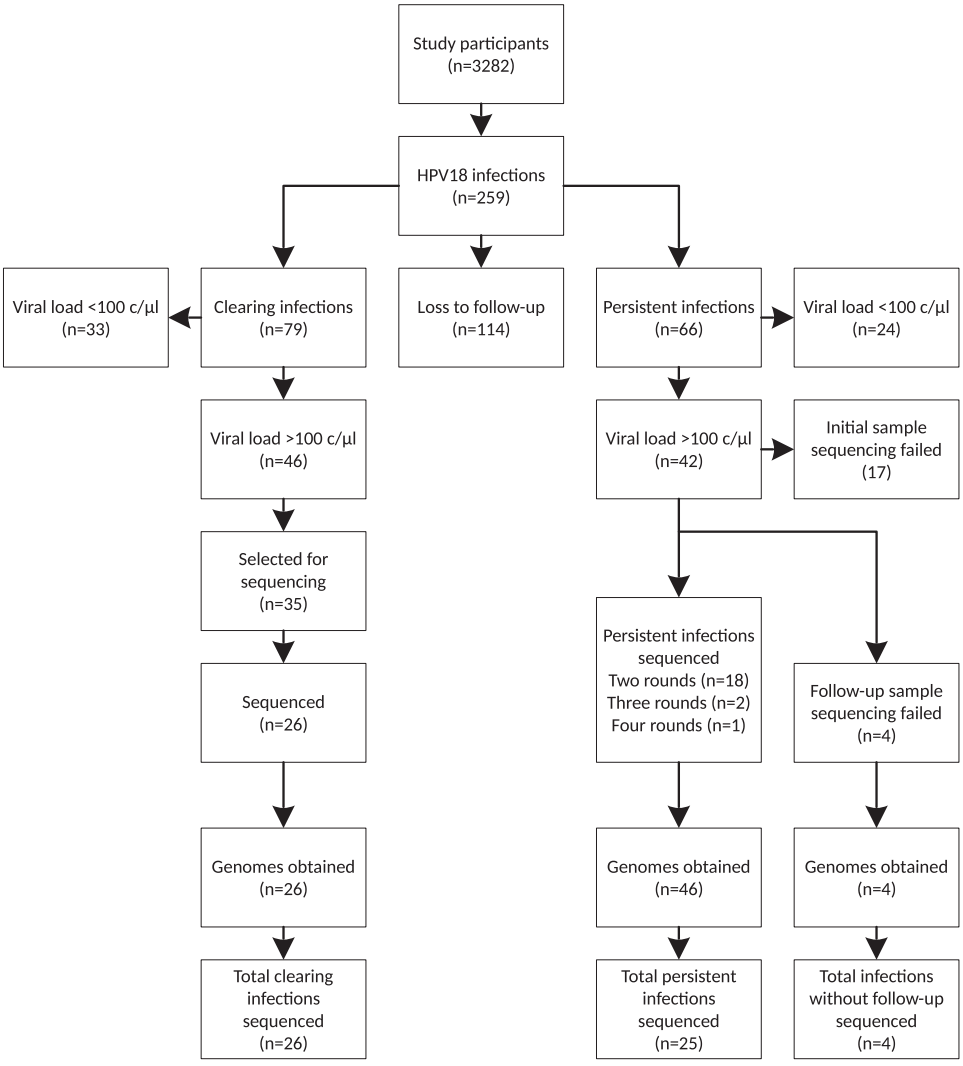


Figure 1. Schematic overview human papillomavirus 18 (HPV18) infections in this study, including loss to follow-up and selections made for persistent and clearing HPV18 infections.

Table 1. General characteristics of study participants with human papillomavirus 18 positive results. Subsets were not found to be significantly different from the complete study population with regard to age, C. trachomatis status and ethnicity ($p > 0.05$).

General Subset Statistics	Infections (n Sequenced/n Total)	
Persistent	25/66	
Clearing	26/79	
Age	Median years (95% CI)	
	Persistent	Clearing
Total dataset	24 (23–25)	25 (23–26)
Sequenced subset	24 (21–25)	23 (20–26)
Unpaired *t*-test	$p = 0.28$	$p = 0.89$
C. trachomatis status	*C. trachomatis* positive (*n*)/Total (*n*)	
	Persistent	Clearing
Total dataset	6/66	8/79
Sequenced subset	3/25	3/26
Fisher's exact test	$p = 0.70$	$p = 1$
European	European (*n*)/Total (*n*)	
	Persistent	Clearing
Total dataset	51/66	55/79
Sequenced subset	18/25	19/26
Fisher's exact test	$p = 0.59$	$p = 0.81$
Mixed	Mixed (*n*)/Total (*n*)	
	Persistent	Clearing
Total dataset	8/66	12/79
Sequenced subset	4/25	4/26
Fisher's exact test	$p = 0.73$	$p = 1.0$
Asian	Asian (*n*)/Total (*n*)	
	Persistent	Clearing
Total dataset	6/66	8/79
Sequenced subset	3/25	2/26
Fisher's exact test	$p = 0.70$	$p = 1.0$
African	African (*n*)/Total (*n*)	
	Persistent	Clearing
Total dataset	0/66	3/79
Sequenced subset	0/25	1/26
Fisher's exact test	$p = 1.0$	$p = 1.0$

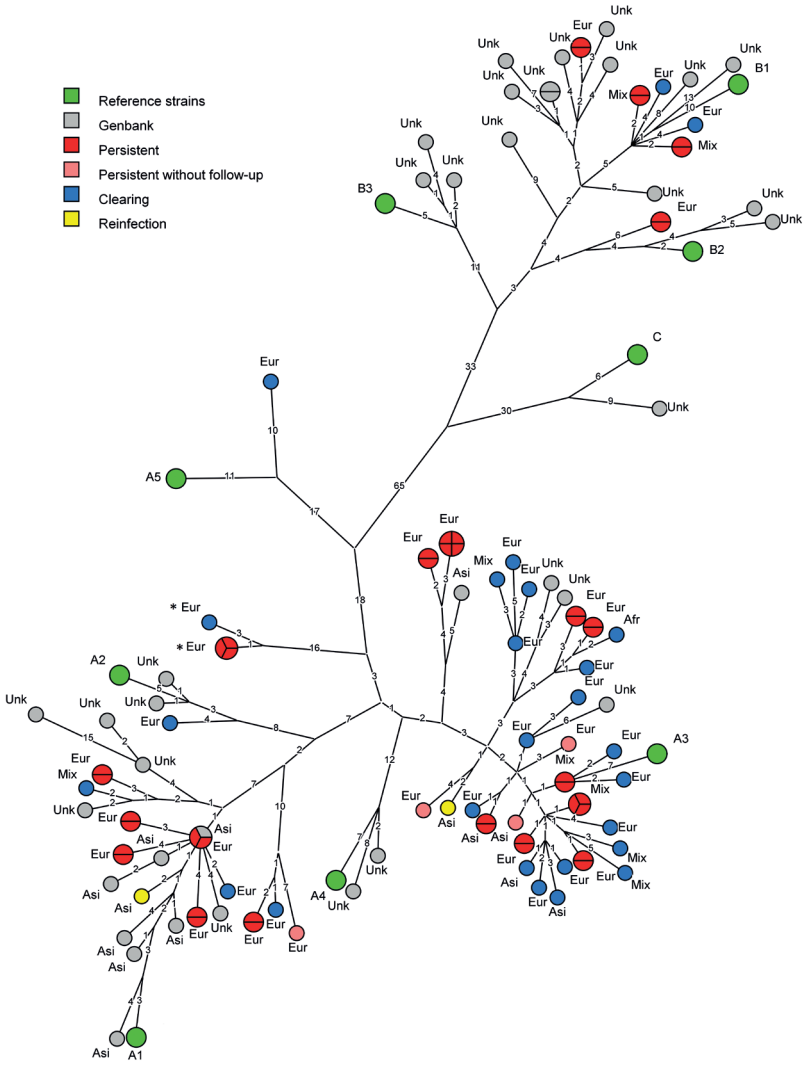


Figure 2. Maximum parsimony phylogenetic tree representing obtained human papillomavirus (HPV) 18 sequence data with corresponding ethnicity and compared to the currently available sequences in Genbank. Numbers on branches indicate the number of nucleotide differences between variants. Each circle represents one sequenced HPV18 variant, with the parts showing how often a variant was found. In green are reference strains for HPV18 according to [4]. Red and pink circles represent sequences obtained from persisting infections with and without sequenced follow-up respectively. Blue circles represent sequences obtained from clearing infections. The yellow circles show an infection where the initial and follow-up samples from one participant cluster differently. Grey circles represent sequences obtained from Genbank. Ethnicity for these sequences was added if available or presented as "Unk" (unknown) if not available. Marked with * are two variants who clustered poorly with any reference strain, but are still closest to A3.

Analysis of Persistent Infections

Follow-up samples of 21 (out of total *n* = 25) persistent infections were sequenced. Of these infections, 18 were sequenced with one round, two with two rounds and one with three rounds of follow-up. In all women with persisting HPV18 infections, the consensus sequences found in follow-up samples were identical to those of the first sample, with one exception. This implies strong conservation of the consensus HPV18 variant within a woman for up to three years. One study participant (sampled 63 weeks apart), initially considered to have a persistent HPV18 infection, actually seemed to have an HPV18 TS reinfection, where the major variant from the initial sample clustered differently from the follow-up sample. This explains that from the 51 women studied here, we identified 52 unique HPV18 sequence variants. The consensus sequences from the apparent reinfection were verified in an independent Illumina sequencing experiment [29]. The Illumina data further showed that the variant present in the follow-up samples, was actually present in the initial samples as well, albeit at a lower ratio than the dominant variant that was identified by Sanger sequencing. The dominant variant that was identified in the initial sample was no longer present in the follow-up sample.

Comparison to Currently Available Data

The genomes sequenced in the current study were compared to the 47 complete HPV18 genomes described previously [14–16, 30, 31] and currently available in NCBI Genbank. The present study adds 51 new unique sequences (Figure 2). Only one sequence in the current dataset has been described previously. Overall the HPV18 whole genome data available in NCBI Genbank is more than doubled by the addition of the sequences described in this study. Especially sublineage A3 is of interest, as only five whole genome sequences have been described for this sublineage. In the Netherlands however, this turns out to be one of the dominant sublineages present in the Netherlands with 33 unique variants, resulting in a large increase of known diversity for this sublineage.

Ethnicity

Of all sequenced HPV18 infections in this study, 72.5% (*n* = 37/51) were found in participants of European ancestry (as proxied using country of birth). Mixed ancestry was reported by 15.7% (*n* = 8/51) of participants, Asian ancestry by 9.8% (*n* = 5/51) of participants and African ancestry by 2.0% (*n* = 1/51) of participants. Distributions of ethnicities stratified for clearing and persistent infections are shown in Table 1 and Figure 2. The distribution of ethnicities of the infections sequenced here was not significantly different from the complete study for any of the represented ethnicities. For this dataset, sequences obtained from participants with non-European ancestry were limited. However, no sublineages seemed to occur preferentially in specific ethnic groups.

SNP Analysis

From the 52 unique HPV18 variants sequenced in this study, 334 SNPs were identified. After exclusion of orphan SNPs not observed in any other participant or follow-up sample, 244 SNPs remained, a large number of which have not been described previously (Figure 3, Table S1). Furthermore, six deletions were found across the various variants. None of the sequences included in this study contained any insertions. Of the 244 non-orphan SNPs, 85 caused amino-acid (AA) changes (Table S2). Three out of six deletion events were found in coding regions of the genome (12 bp E1, 6 bp E2/E4 and 15 bp L2). None of the found deletions caused frameshifts or AA changes, including the deletion found in the shared open-reading frame (ORF) for E2 and E4. The other three deletions were identified in the intergenic region between E2 and E5 ORFs (20 bp) and the upstream regulatory region (URR) (26 bp and 7 bp). The E2 deletion and the 7 bp URR were linked for this study. Six participants were found with both deletions and none with either one of the deletions. In addition five of these six study participants also had the 20 bp deletion between the E2 and E5 ORFs, suggesting another link, albeit not exclusive. This deletion was not identified on its own in any of the other study participants. The 12 bp E1, 15 bp L2 and 26 bp URR deletions have not been described previously, although they were only identified in one study participant each.

An SNP comparison was made between sequences obtained from persisting and clearing HPV18 infections. Upon exclusion of the variants identified in the reinfection event and the variants for which the follow-up samples could not be sequenced, SNPs were compared between 20 persistent infections and 26 clearing infections. No individual SNP was found to be significantly different between the two groups. A sliding window analysis of genome diversity showed no large differences between the two groups (Figure 4).

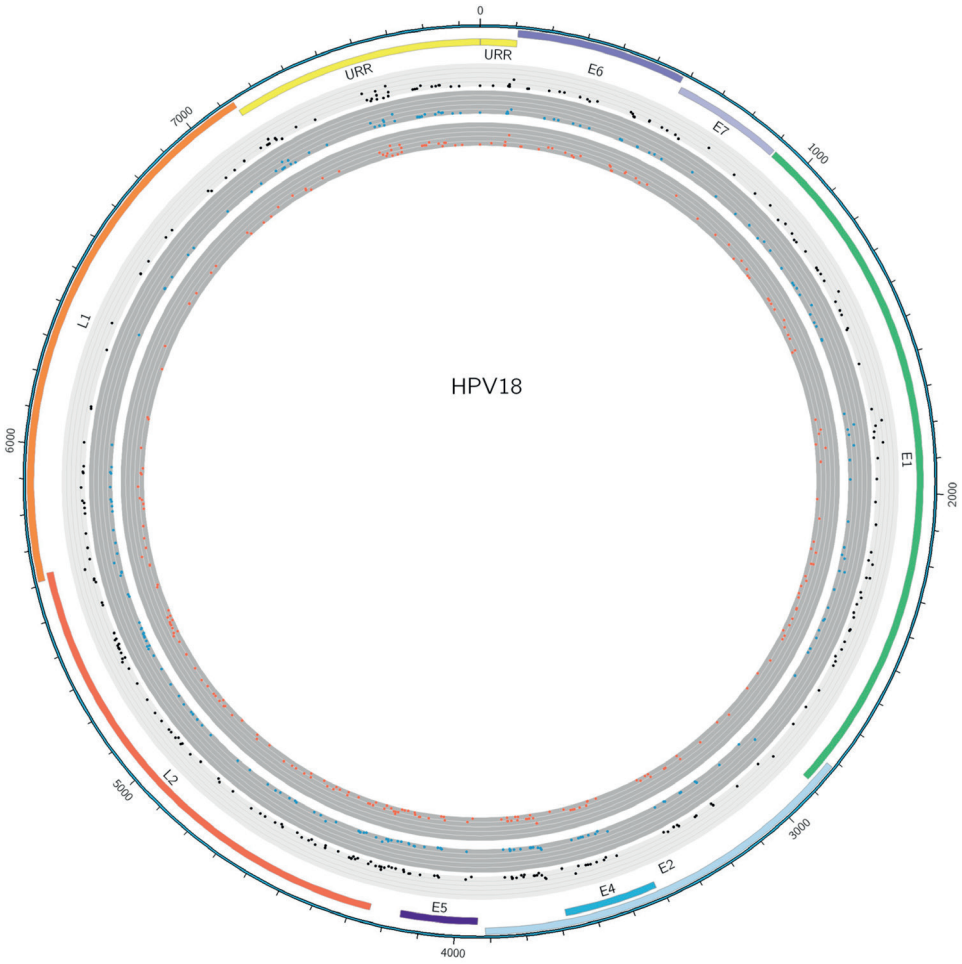


Figure 3. Circular plot of non-orphan single-nucleotide polymorphisms (SNPs) on the human papillomavirus (HPV) 18 genome. Each dot denotes a variable site within the dataset. Black dots show total SNPs occurring in this study. Blue dots were variations identified in clearing infections, while reds dots were variations identified in persisting infections. Total DNA variations are shown on the light grey circles. Height of dots on the respective circles shows how often the variation occurs in the dataset.

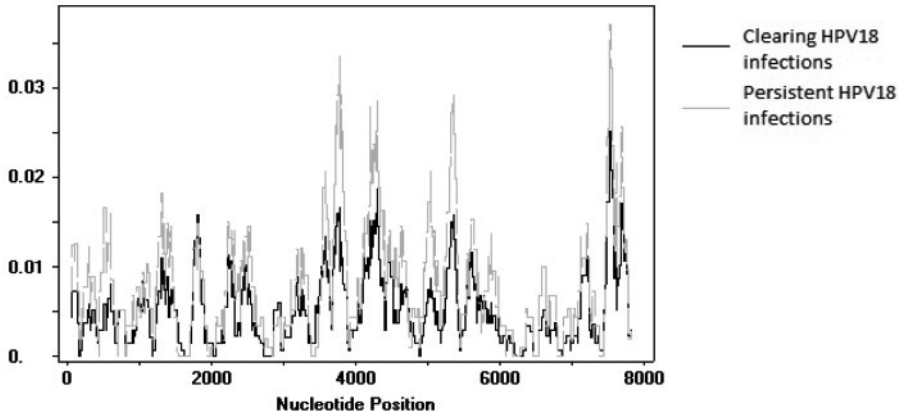


Figure 4. Plot of nucleotide diversity (pi) of human papillomavirus (HPV) 18 sequences obtained in this study. The black line shows data from clearing infections (n = 25), while the grey line shows data from persistent infections (n = 20).

Discussion

In this study, we describe the remarkably large diversity of HPV18 variants circulating in a Dutch cohort study based on whole genome sequencing data. Due to one HPV18 infection with two different consensus variants at follow-up sampling moments, we actually identified 52 unique variants from 51 study participants. Despite this diverse population in study participants, we find that in persistent infections, the same consensus sequence remains completely conserved in up to three years of follow-up.

With this study, the number of published complete HPV18 genome sequences has been doubled. Of the 52 unique sequences found in our study, 51 are unique in the NCBI Genbank database for whole genome HPV18 sequences. To date only three studies have described and compared multiple HPV18 whole genome sequences, highlighting the necessity of HPV18 population studies. These three studies have included samples from South-America, Africa and South-east Asia. European studies are scarce and to our knowledge we are the first to describe whole genome HPV18 diversity for a European country. Interestingly, phylogenetic analysis shows the variant diversity partially matches the variants described to date, but in this study there is a heavy emphasis on sublineage A3 variants, which is underrepresented in other studies. This suggests that in the Netherlands, and possibly Europe, there is a different distribution of sublineages than in other parts of the world. In both phylogenetic analyses performed in this study, persisting and clearing infections seem to be distributed differently in the A1 and A3 clades, however, the distribution of clearing and persistent infections between these clades was not statistically significant. This may be due to the relatively small sample size and it would be of interest to see if these differences in distribution also occur in larger datasets.

It should be noted that during initial amplification, seventeen persisting and nine clearing infections failed to amplify. A further four persistent infections were successfully sequenced in the initial round, but failed in the follow-up sample. This might suggest a PCR bias in the sequenced group and a possible underrepresentation of variants that could not be amplified by our PCR assay. This explanation seems unlikely because the primers used for this study, as they were designed against conserved areas of the HPV18 genome and variants from six different sublineages were successfully amplified.

Considering the single occurrence of an HPV18 infection with two different major variants in follow-up samples that we found from 21 study participants with persistent HPV18 infections, we can conclude that conventional genotyping leads to reliable identification of persistent infections. However, in the context of vaccine studies, a single persistent infection could affect measured vaccine efficacy or efficiency in a negative manner, while this might actually be a repeat incident infection after sequencing. In such specific cases, it could be warranted to confirm if it is indeed a persistent infection, or actually a type-specific reinfection. In fact, the next-generation sequencing (NGS) confirmation of apparent reinfection [29], suggested that at the initial sampling moment, the variant from the follow-up sample was already present at a low ratio. Upon reinspection of the Sanger data, secondary peaks could be distinguished at low signal for the variable sites identified by NGS. This suggests that this infection meets both the criteria for a persisting (HPV18 positive with HPV18 positive follow-up) and a clearing infection (HPV18 positive, with HPV18 negative follow-up), based on the variant analyzed. In the follow-up sample, the initial variant was not identified from the NGS experiment. This suggests that our Sanger sequencing results were reliable in assessing persistent infections where the same consensus sequence is identified in follow-up samples, but not in the case of reinfections, where a cocktail of variants might be present and variant-specific clearance may occur.

The data obtained in this study identifies a large number of new SNPs. In addition, six deletions were found, of which the coding one on E2/E4 and the non-coding ones on the E2 E5 intergenic region and the URR (7 bp) have been described previously [14–16]. Three other deletions which have not been described previously were identified once each in this study population. This could suggest a possible artifact; however, the deletions on E1 and the URR (26 bp) were identified from persistent infections. The same deletions were found in three rounds for the E1 deletion and in two rounds for the 26 bp URR deletions which strongly argues for true signs of diversity and not diagnostic artifacts. The deletion found in L2 was only identified once from a clearing infection. This deletion, along with orphan SNPs identified only once in this dataset, could not be confirmed due to scarcity of the material and therefore we cannot exclude that these represent possible PCR artifacts.

The genomes described in this study were obtained through Sanger sequencing. In the current era, next-generation sequencing is rapidly emerging as the go-to method for sequencing. It generates far more data at a higher sensitivity than Sanger sequencing, thereby allowing for the identification of intra-patient HPV co-infections and even HPV type-specific variant co-infections [9]. Such findings are beyond the resolution of Sanger sequencing. Despite this, Sanger remains

the golden standard in sequencing and, for our rationale, an adequate, cheap technique to use. The robustness of Sanger sequencing is shown in this study by the consistent identification of the same consensus sequence through time in persistent infections.

In addition, we have attempted to compare sequences obtained from persistent and clearing infections. For HPV16, it has been shown that E7 conservation is higher in infections that eventually progress to CIN3+ [10]. In this study, we do observe strong conservation of E7 for HPV18, but the number of infections sequenced in this study is too small to identify any differences between sequences obtained from persistent or clearing infections. To gain insight into the role of the HPV18 genome on infection duration, large-scale NGS studies should be performed.

Previous studies have been ambiguous in describing an association between HPV18 (sub) lineages and specific ethnic groups with regards to persistent infections and the development of cervical cancer [8, 27, 32, 33]. The dataset in this study is too limited in size to identify clear links between the occurrence of (persistent) HPV18 variants in population subgroups.

Despite the limited study size, we have shown the natural occurrence of a large pool of HPV18 variants in a Dutch population of young women. Our findings stress the importance of identifying the different properties associated with various variants and (sub)lineages for HPV18. When comparing clearing and persistent infections, no specific SNPs were identified as strongly affecting the outcome of infection, although the size of the dataset used here could only identify very strong-acting effects.

Supplementary Materials

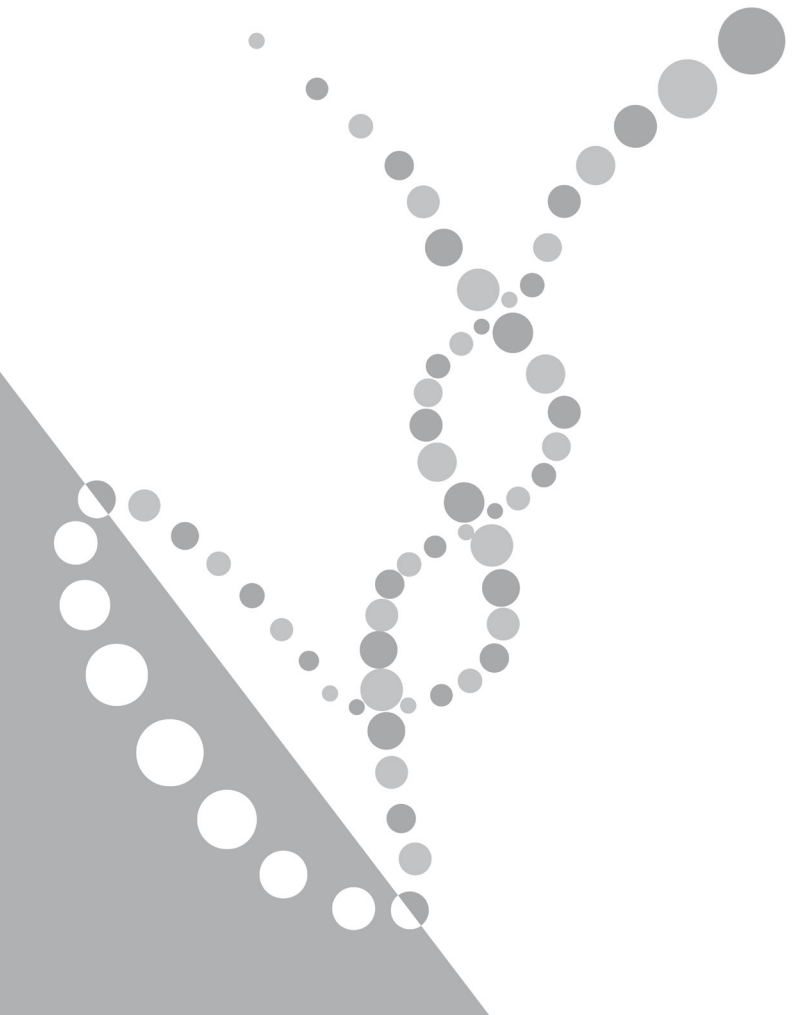
The following are available online at www.mdpi.com/link, Figure S1: Maximum likelihood tree of sequences obtained in this study, Table S1: Nucleotide alignment of sequences obtained in this study, aligned to reference strain AY262282, Table S2: Amino-acid alignment of sequences obtained in this study, aligned to reference strain AY262282.

References

1. Walboomers, J.M., et al., *Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.* J Pathol, 1999. **189**(1): p. 12–9.
2. Humans, I.W.G.o.t.E.o.C.R.t., *Biological agents. Volume 100 B. A review of human carcinogens.* IARC Monogr Eval Carcinog Risks Hum, 2012. **100**(Pt B): p. 1–441.
3. de Sanjose, S., et al., *Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study.* Lancet Oncol, 2010. **11**(11): p. 1048–56.
4. Burk, R.D., A. Harari, and Z. Chen, *Human papillomavirus genome variants.* Virology, 2013. **445**(1–2): p. 232–43.
5. Xi, L.F., et al., *Risk for high-grade cervical intraepithelial neoplasia associated with variants of human papillomavirus types 16 and 18.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(1): p. 4–10.
6. Xi, L.F., et al., *Lineages of oncogenic human papillomavirus types other than type 16 and 18 and risk for cervical intraepithelial neoplasia.* J Natl Cancer Inst, 2014. **106**(10).
7. Mirabello, L., et al., *HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women.* J Natl Cancer Inst, 2016. **108**(9).
8. Chen, A.A., et al., *Human Papillomavirus 18 Genetic Variation and Cervical Cancer Risk Worldwide.* J Virol, 2015. **89**(20): p. 10680–7.
9. Cullen, M., et al., *Deep sequencing of HPV16 genomes: A new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection.* Papillomavirus Research, 2015.
10. Mirabello, L., et al., *HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis.* Cell, 2017. **170**(6): p. 1164–1174 e6.
11. Xi, L.F., et al., *Variant-specific persistence of infections with human papillomavirus Types 31, 33, 45, 56 and 58 and risk of cervical intraepithelial neoplasia.* Int J Cancer, 2016. **139**(5): p. 1098–105.
12. van Belkum, A., et al., *Genotyping human papillomavirus type 16 isolates from persistently infected promiscuous individuals and cervical neoplasia patients.* J Clin Microbiol, 1995. **33**(11): p. 2957–62.
13. van der Weele, P., C. Meijer, and A.J. King, *Whole-Genome Sequencing and Variant Analysis of Human Papillomavirus 16 Infections.* J Virol, 2017. **91**(19).
14. Chen, Z., et al., *Evolutionary dynamics of variant genomes of human papillomavirus types 18, 45, and 97.* J Virol, 2009. **83**(3): p. 1443–55.
15. Lurchachaiwong, W., et al., *Whole-genome sequence analysis of human papillomavirus type 18 from infected Thai women.* Intervirology, 2010. **53**(3): p. 161–6.
16. Chen, Z., et al., *Evolution and taxonomic classification of alphapapillomavirus 7 complete genomes: HPV18, HPV39, HPV45, HPV59, HPV68 and HPV70.* PLoS One, 2013. **8**(8): p. e72565.
17. Apter, D., et al., *Efficacy of human papillomavirus 16 and 18 (HPV-16/18) AS04-adjuvanted vaccine against cervical infection and precancer in young women: final event-driven analysis of the randomized, double-blind PATRICIA trial.* Clin Vaccine Immunol, 2015. **22**(4): p. 361–73.
18. Joura, E.A., et al., *A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women.* N Engl J Med, 2015. **372**(8): p. 711–23.
19. van den Broek, I.V., et al., *Evaluation design of a systematic, selective, internet-based, Chlamydia screening implementation in the Netherlands, 2008–2010: implications of first results for the analysis.* BMC Infect Dis, 2010. **10**: p. 89.

20. van den Broek, I.V., et al., *Effectiveness of yearly, register based screening for chlamydia in the Netherlands: controlled trial with randomised stepped wedge implementation.* BMJ, 2012. **345**: p. e4316.
21. Mollers, M., et al., *Prevalence, incidence and persistence of genital HPV infections in a large cohort of sexually active young women in the Netherlands.* Vaccine, 2013. **31**(2): p. 394–401.
22. Woestenberg, P.J., et al., *Comparison of STI-related consultations among ethnic groups in the Netherlands: an epidemiologic study using electronic records from general practices.* BMC Fam Pract, 2015. **16**: p. 70.
23. Kleter, B., et al., *Development and clinical evaluation of a highly sensitive PCR-reverse hybridization line probe assay for detection and identification of anogenital human papillomavirus.* J Clin Microbiol, 1999. **37**(8): p. 2508–17.
24. Kleter, B., et al., *Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses.* Am J Pathol, 1998. **153**(6): p. 1731–9.
25. van der Weele, P., et al., *Correlation between viral load, multiplicity of infection, and persistence of HPV16 and HPV18 infection in a Dutch cohort of young women.* J Clin Virol, 2016. **83**: p. 6–11.
26. Arroyo, S.L., et al., *Human papillomavirus (HPV) genotype 18 variants in patients with clinical manifestations of HPV related infections in Bilbao, Spain.* Virol J, 2012. **9**: p. 258.
27. King, A.J., et al., *Genetic Diversity in the Major Capsid L1 Protein of HPV-16 and HPV-18 in the Netherlands.* PLoS One, 2016. **11**(4): p. e0152782.
28. Cole, S.T. and O. Danos, *Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products.* J Mol Biol, 1987. **193**(4): p. 599–608.
29. Van der Weele, P., *National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven, the Netherlands. Illumina data of a HPV18 apparent reinfection event.* 2016.
30. Siqueira, J.D., et al., *Identification of novel human papillomavirus lineages and sublineages in HIV/HPV-coinfected pregnant women by next-generation sequencing.* Virology, 2016. **493**: p. 202–8.
31. Oliveira, G.R., et al., *Characterisation of complete high- and low-risk human papillomavirus genomes isolated from cervical specimens in southern Brazil.* Mem Inst Oswaldo Cruz, 2017. **112**(10): p. 728–731.
32. Xi, L.F., et al., *Human papillomavirus type 16 and 18 variants: race-related distribution and persistence.* J Natl Cancer Inst, 2006. **98**(15): p. 1045–52.
33. Lizano, M., et al., *Distribution of HPV16 and 18 intratypic variants in normal cytology, intraepithelial lesions, and cervical cancer in a Mexican population.* Gynecol Oncol, 2006. **102**(2): p. 230–5.

Chapter 6



HPV16 whole genome minority variants in persistent infections from young Dutch women

Sonja Lagström[*,a,b,c], Pascal van der Weele[*,d,e], Trine Ballestad Rounge[b], Irene Kraus Christiansen[a,f], Audrey J. King[d] and Ole Herman Ambur[g]

[*]S. L. and P. v. d. W. have contributed equally to this work
[a]Department of Microbiology and Infection Control,
Akershus University Hospital, Lørenskog, Norway.
[b]Department of Research, Cancer Registry of Norway, Oslo, Norway.
[c]Institute of Clinical Medicine, University of Oslo, Oslo, Norway.
[d]National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Research, Diagnostics and Screening, Bilthoven, the Netherlands
[e]Vrije Universiteit-University Medical Center (VUmc),
Department of Pathology, Amsterdam, the Netherlands
[f]Department of Clinical Molecular Biology (EpiGen), Division of Medicine, Akershus University Hospital and Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[g]Faculty of Health Sciences, OsloMet–Oslo Metropolitan University, Oslo, Norway

J Clin Virol. 2019 Aug 12;119:24-30

Abstract

Background Chronic infections by one of the oncogenic human papillomaviruses (HPVs) are responsible for near 5% of the global cancer burden and HPV16 is the type most often found in cancers. HPV genomes display unexpected levels of variation when deep-sequenced. Minor nucleotide variations (MNVs) may reveal HPV genomic instability and HPV-related carcinogenic transformation of host cells.

Objectives The objective of this study was to investigate HPV16 genome variation at the minor variant level on persisting HPV16 cervical infections from a population of young Dutch women.

Study design 15 HPV16 infections were sequenced using a whole-HPV genome deep sequencing protocol (TaME-seq). One infection was followed over a three-year period, eight were followed over a two-year period, three were followed over a one-year period and three infections had a single sampling point.

Results and conclusions Using a 1% variant frequency cutoff, we find on average 48 MNVs per HPV16 genome and 1717 MNVs in total when sequencing coverage was >100×. We find the transition mutation T>C to be the most common, in contrast to other studies detecting APOBEC-related C>T mutation profiles in pre-cancerous and cancer samples. Our results suggest that the relative mutagenic footprint of HPV16 genomes may differ between the infections in this study and transforming lesions. In addition, we identify a number of MNVs that have previously been associated with higher incidence of high-grade lesions (CIN3+) in a population study. These findings may provide a starting point for future studies exploring causality between emerging HPV minor genomic variants and cancer development.

Highlights

- HPV16 genomes display variation in individual cervical samples
- Emerging genomic minor nucleotide variations in persistent HPV16 infections
- Frequent T>C mutational profiles were identified
- Minor nucleotide variations previously associated with high grade lesions (CIN3+) detected

Introduction

Human papillomavirus (HPV) is the most common sexually transmitted infection worldwide [1] and persistent infection with an oncogenic HPV type is required, but not sufficient, for the development of cervical cancer [2]. Although most HPV infections clear naturally within 12–18 months [3], a subset may persist, potentially progressing to cervical intraepithelial lesions of varying degrees (CIN1–3) and invasive cervical cancer. HPV is a double stranded DNA virus that uses host replication machinery and has co-diverged with humans to constitute highly conserved genotypes [4, 5]. Within HPV types, distinction is made between lineages (1–10% whole genome genetic difference), sublineages (0.5–1.0%) and variants (< 0.5%) [6]. Lineages and sublineages of HPV have been associated with differential risks for disease outcomes [7, 8]. In addition, recent studies have shown that HPV exhibits large variation, both at the population level and within its human host despite the strongly conserved genome [9–14]. Currently, limited information is available explaining the origin of this diversity.

Deep sequencing of HPV genomes has revealed the presence of minor nucleotide variations (MNVs). These polymorphic sites show one or multiple different nucleotides in addition to the consensus or majority nucleotide [15]. Such MNVs can only be reliably detected by means of high-resolution sequencing. HPV is considered to evolve slowly due to the high fidelity of its human host replication machinery [4, 16]. However, humans also encode several low-fidelity polymerases, some of which are upregulated in early stages of HPV16 infection [17]. These polymerases are often recruited for DNA repair by means of non-homologous end-joining [18]. The HPV life cycle involves two separate rounds of replication. An initial round in proliferating cells at the basal layer of the stratified epithelium yielding 10–100 copies of the viral genome per cell, and another productive round, in differentiated cells at the suprabasal level, resulting in thousands of viral copies per cell [19]. Several DNA repair pathways are required for productive HPV replication, yet information relating to their influence on generating mutations at the minor variant level is lacking. In addition, viral mutation rates can be affected by, sequence context, template secondary structure, the cellular microenvironment and several other factors relating to replication, post-replicative corrections and DNA repair [20]. A known source of mutations in HPV genomes is apolipoprotein B mRNA editing enzyme (APOBEC) activity, which is part of the host innate immune response against viruses [21]. APOBEC enzymes induce genetic change by converting cytidine to uridine, which may base pair with adenosine, causing C > T substitution mutants after replication. APOBEC-related changes have been identified in cervical cancer patient genomes [22]. Additionally, the HPV genome is itself susceptible to APOBEC editing [12, 23]. HPV oncoproteins E6 and E7 upregulate the expression of APOBEC3A and APOBEC3B [24, 25]. In turn, APOBEC3B activity is upregulated in cancer tissues [26, 27]. Interestingly, conservation of the HPV *E7* gene, through a lack of APOBEC-related editing, was shown to be essential for the development of cervical cancer in a population study [12]. Despite these findings, APOBEC activity in HPV infections in young women remains largely uncharacterized.

In this study, we aim to identify intra-sample MNVs in HPV16 infections from young women and monitor changes over time. To this end, we use TaME-seq for sequencing [28]. TaME-seq adapts tagmentation-assisted (enzymatic cleaving and tagging of double-stranded DNA) library preparation by replacing one of the generic sequencing primers with a cocktail of 52 HPV specific primers. Reactions are performed separately for forward and reverse sequencing products, replacing the forward generic primer with a HPV specific one and vice versa. This multiplex PCR enrichment approach results in a higher yield of HPV specific sequencing data. Here, we apply TaME-seq, to a longitudinal retrospective cohort study [29].

Materials and methods

Sample selection

Vaginal self-swabs were obtained from the *Chlamydia trachomatis* Screening and Implementation (CSI) study. Recruitment criteria, methods and additional consent for HPV testing have been described previously [29–31]. Cytology was not performed on these samples, but considering the age of study participants (16–29 years old), the identified infections are likely benign. Participants supplied up to four samples over time. For this study, the median interval between sampling moments was 48 weeks (95% CI: 46–51 weeks; min: 17, max: 63 weeks). Total DNA from 200 µL of sample was isolated using the MagnaPure96 platform (Total Nucleic Acid Isolation Kit, Roche Diagnostics) according to the manufacturer's protocol. Isolated material was eluted in 100 µL and subsequently genotyped via the SPF10-DEIA-LiPA25 platform (DDL Diagnostics) [32, 33]. Viral load of HPV16 positive samples was quantified via type-specific qPCR [34]. Infections were selected if they were HPV16 positive during at least three subsequent follow-up moments, preferably with no other HPV genotypes present (Figure 2).

Library preparation and sequencing

Library preparation was performed using TaME-seq [28]. Briefly, each sample was tagmented using the Nextera DNA library prep kit (Illumina, Inc., San Diego, CA) and subsequently amplified in two separate reactions. Amplification occurred by multiplex PCR using pools of 27 forward (F) and 25 reverse (R) HPV16 primers in combination with i7 and i5 index primers [35] from the Nextera index kit (Illumina, Inc., San Diego, CA). Libraries from all samples were sequenced on the Illumina MiSeq and HiSeq2500 platform as 151 bp paired-end reads with two 8 bp index reads.

Sequence alignment and nucleotide variant calling

Sequence data was analyzed using an in-house bioinformatics pipeline [28]. Reads were mapped to the human genome (GRCh38/hg38) and HPV16 reference genome (GI:333031 HPV16REF.1) [36], using HISAT2 (v2.1.0) [37]. Consensus sequences were extracted using samtools (v1.8) mpile-

up (-E -d 200000 -L 200000), bcftools (v1.6) (call -c —ploidy 1) and vcfutils.pl. Consensus sequences were compared to Sanger data from a previous study [10] using MUSCLE (v.3.8.1551) to align sequences, IQtree (v1.5.5) to infer maximum likelihood phylogeny and FigTree (v1.4.3) to visualize the alignment. Mapped nucleotide counts over HPV reference genomes and average mapping quality values of each nucleotide were retrieved from BAM files. Variant calling was performed using an in-house R (v3.4.4) script (Figure 1). In each sample, nucleotides called ≤2 times in each genomic position or with mean Phred score of <30 were removed. From either reaction, results with coverage <100× were filtered out. F and R nucleotide counts were pooled per sample and major and minor variant frequencies were calculated per position. Samples were excluded if <45% of the genome was covered ≥100×. Variants were called if variant frequency was >1%. If F and R reactions from the same sample showed discordant variants, the reaction with higher coverage was chosen for total variant calling. Genomic locations of MNVs were mapped and major to minor variant mutations were classified as synonymous or non-synonymous in each infection. In addition, MNVs appearing consecutively in follow-up samples from the same infection were identified. Selected samples with a high read count (>1,000,000) mapped to HPV16, were downsampled randomly to 100.000 reads to rule out possible effects of excessively high sequencing coverage on variant calling.

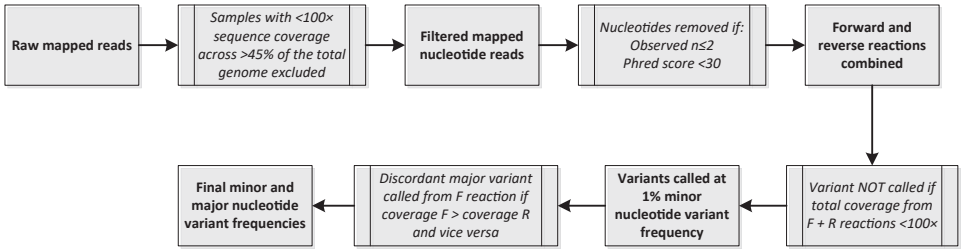


Figure 1: Schematic representation of the nucleotide variant calling.

Mutational signature analysis

All observed nucleotide substitutions were classified into the six base substitutions, C>A (G>T), C>G (G>C), C>T (G>A), T>A (A>T), T>C (A>G), and T>G (A>C) substitutions, and then into 96 trinucleotide substitution types that include information on the bases immediately 5' and 3' of the mutated base. Analysis was performed using an in-house R (v3.4.4) script. A region frequently subject to insertions / deletions (indels) was identified in the non-coding region (NCR) at positions 4184 and 4185. At these positions, small indels in the sequenced genomes often resulted in mapping errors. Consequentially, 18 T>A and 2 T>G mutations in these two positions have been removed from the present analysis.

Data availability

The data obtained in this study was deposited in ENA under project number (will be added when available).

Results

Mean sequencing coverage and viral load

In total, 59 samples from persistent HPV16 infections were processed using TaME-seq and 61% (36/59) had >45% genome covered by minimum 100× (Table S1), which was the criterion for further analysis. The remaining 36 samples originated from 15 infections (Figure 2). The mean sequencing coverage per sample ranged from 653 to 399,653 reads (Table S1). Samples had varying HPV16 viral load, which correlated strongly with the per sample mean sequencing coverage (Figure 3, Pearson correlation coefficient 0.89).

Of the samples with a high viral load (>1500 copies/µL; n=30), 29 could be included in downstream analyses. Of the samples with a lower viral load (<1500 copies/µL; n=29), only seven could be included, bringing the total sample number included in downstream analyses to 36.

Comparison of NGS data to previous Sanger results

Sanger data from a previous study was available for 29 out of 36 samples [10]. Consensus sequences obtained in the present study were compared to those previously described [10]. The alignment of Sanger and NGS results overlaps, suggesting high concordance between datasets (Figure S1).

HPV16 minor nucleotide variations

A total of 1717 HPV16 MNVs (variant frequency >1% and coverage ≥100×) were detected in the 36 samples (Table 1; Figure 4; Table S2), with 15 to 82 different variants per sample (average 48.3 variants per genome, Table S2). Variant frequency ranged from 1% to 49.6%. No significant correlation was found between the number of variable sites detected and mean or median sequencing coverage (Pearson correlation coefficient: r=-0.41). We note however that the sample (545351–3) with the by far highest mean coverage (399653) and viral load reports the lowest number of variable sites (n=15) (Table S1 and S2). The two samples (340223–1 and 407612–1) with the lowest mean coverage, report the mean (48) or below (32) number of variable sites (Table S1 and S2). Of all variants, 85.3% (1465/1717) had a frequency of <5%. Non-synonymous and synonymous MNVs were analyzed and are summarized in Table 2.

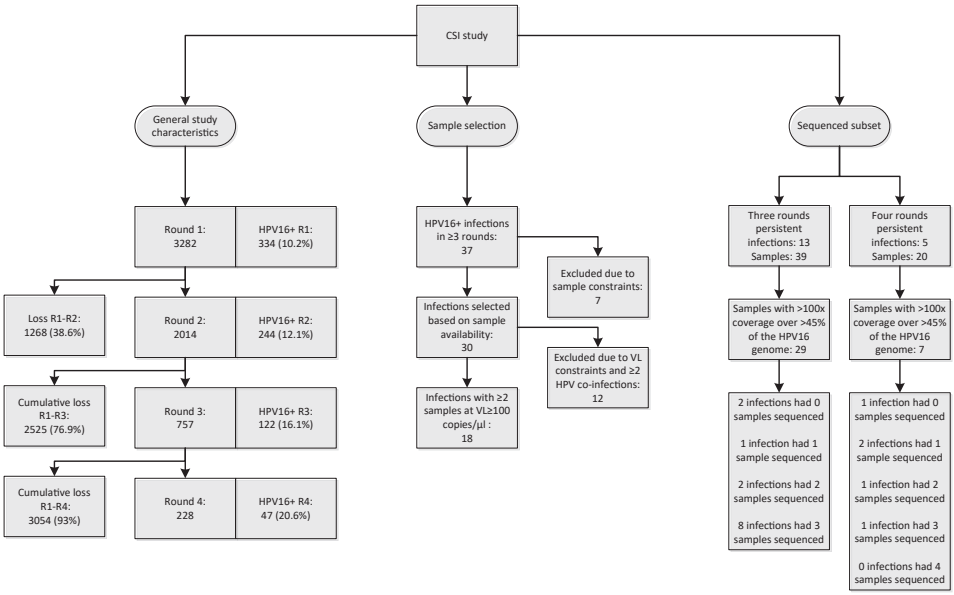


Figure 2: Study flowchart describing selected samples and sequencing outcome. VL = viral load

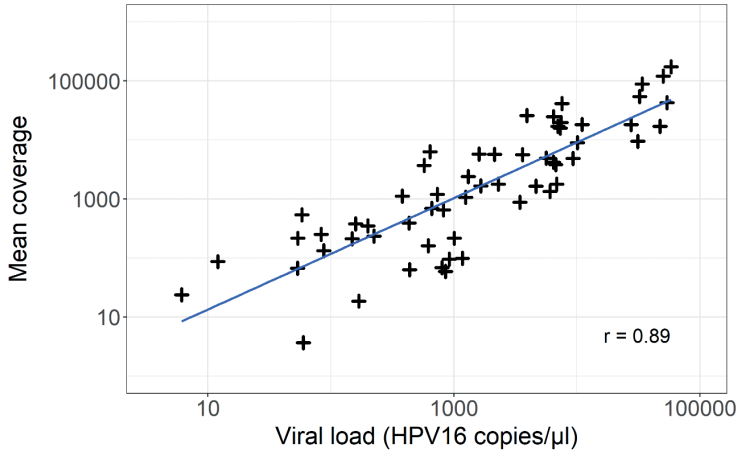


Figure 3: Correlation between mean sequencing coverage and viral load (HPV16 copies/µL) in each sample.

Table 1: Composition of minor nucleotide variations (MNVs) identified in this study. Percentages of total (%) MNVs were identified from a single reaction of either the forward (F) or reverse (R) sequencing reaction (only F or R, coverage >100x) or from both sequencing reactions (F and R, coverage >100×). MNVs identified in regions where F and R overlapped were compared to major nucleotides and scored if they matched (same major, same MNVs from both F and R) or mismatched (same major, different MNV from both F and R). Finally the number of MNVs detected repeatedly in follow-up samples is shown.

MNV calling/mutation type	T>C	T>A	T>G	C>T	C>A	C>G	Total
Total MNVs	1146 (67%)	56	68	404 (23%)	26	17	1717
MNVs with coverage >100× for either F or R	610 (68%)	36	38	190 (21%)	13	8	895
MNVs with coverage >100× for both F and R	536 (65%)	20	30	214 (26%)	13	9	822
Same major and different MNV F and R	398 (63%)	16	22	175 (28%)	8	6	625
Same major and same MNV F and R	135 (73%)	2	6	35 (19%)	4	2	184
Consecutive detection of same MNV	31 (44%)	3	6	24 (34%)	4	2	70

In order to explore unusual mutational patterns in any gene region, the number of synonymous and non-synonymous MNVs was mapped against the consensus sequence of each infection (Table 2). On average there were 1,67 times (STDEV ±0,19) more non-synonymous than synonymous mutations. No genomic region could be singled out as notably different from other regions.

The total number of MNVs observed in each gene region varied considerably (Table 2), but correlated well with gene length (Pearson correlation coefficient: 0.98; Figure 5). The *L2* gene showed a lower than expected amount of variation, although sequence coverage was low around genome positions 4800–5000 bp. Overall, the majority of MNVs found (90%, n = 1550/1717), were caused by transition events (Table 1). Transversion mutations were detected in 10% of cases. The most common MNV was T > C (A > G; 67%, n = 1146/1717) followed by C > T (G > A; 24%, n = 404/1717) (Table 1; Figure 6). The overall T > C mutation ratio was 67%. In comparison, the T > C ratios identified from either F or R sequencing reactions or both sequencing reactions together were 68% and 65% respectively (Table 1). When MNVs were detected in regions where F and R reactions overlapped, the T > C ratio was 63% when either the F/R reactions identified a MNV over the set threshold (same major different minor) and 73% when both F/R reactions made the same MNV call (same major and same MNV) (Table 1).

Consecutive samples collected at one-year intervals from the same infection generally showed different MNVs over time. However, 35 MNVs across the HPV16 genome were recaptured in one of the follow-up samples of eleven different infections (Table S3) amounting to 4% (70/1717) of the total MNVs. Furthermore, the T > C mutation ratio drops to 44% in this subset relative to the overall ratio. The T > C MNVs were the most prevalent in all but one sample collected at the third sampling point (444086–3), where the C > T minor variants were dominant (Figure S2, S3).Moreover, 45 MNVs were detected at 21 polymorphic sites previously associated with CIN3 + (Table S4) [12]. Of these, the two polymorphisms most frequently found were seven in position 3410 in the E2/E4 gene and six in position 4042 in the E5 gene.

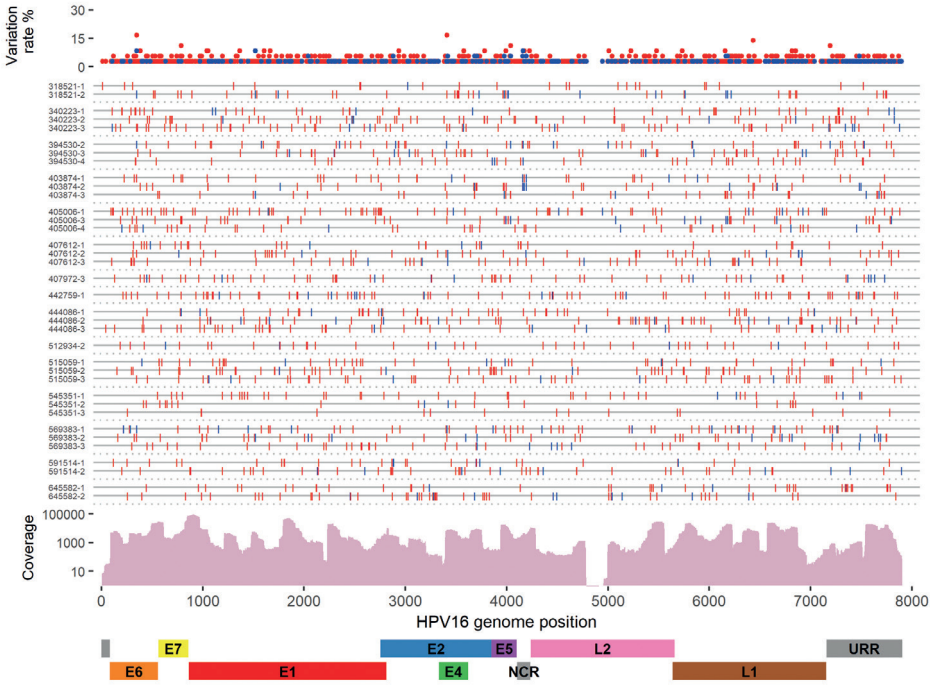


Figure 4: Variable sites (n = 1717) and mean sequencing coverage in the 36 samples from 15 individuals. Variation rate (top) shows the amount of samples (in %) carrying a minor nucleotide variant in each position. Each horizontal line represents an individual sample, which is named according to case number and sample number (1–4) indicating the sample collection time point. Samples from the same infection are clustered and separated from others by dashed lines. Variable positions with variant frequency of ≤5% are marked with red and variable positions with variant frequency >5% is marked with blue. Mean sequencing coverage is shown across the HPV16 genome. The location of early (*E1*, *E2*, *E4–7*), late (*L1*, *L2*) genes, URR and NCR is indicated below the HPV16 genomic positions.

Table 2: Minor nucleotide variants per HPV16 gene/genome region in the 36 HPV16 samples included in the analysis. Where applicable, MNVs are sorted by effect on coding sequence relative to the major nucleotide variant of each infection.

Gene	Length (bp)	Total number (n) of minor nucleotide variations			
		All (%)	Synonymous (%)	Non-synonymous (%)	Nonsense (%)
E6	477	116 (24.3)	46 (9.6)	66 (13.8)	4 (0.8)
E7	297	65 (21.9)	26 (8.8)	39 (13.1)	0
E1	1950	439 (22.5)	143 (7.3)	282 (14.5)	14 (0.7)
E2	1098	248 (22.6)	90 (8.2)	155 (14.1)	3 (0.3)
E4	288	65 (22.6)	25 (8.7)	40 (13.9)	0
E5	252	68 (27.0)	23 (9.1)	44 (17.5)	1 (0.4)
L2	1422	233 (16.4)	91 (6.4)	142 (10.0)	0
L1	1518	350 (23.1)	133 (8.8)	210 (13.8)	7 (0.5)
URR	832	180 (21.6)	-	-	-
SUM		1801*	577	978	29

*Since certain genes overlap, 84 MNVs are reported more than once.

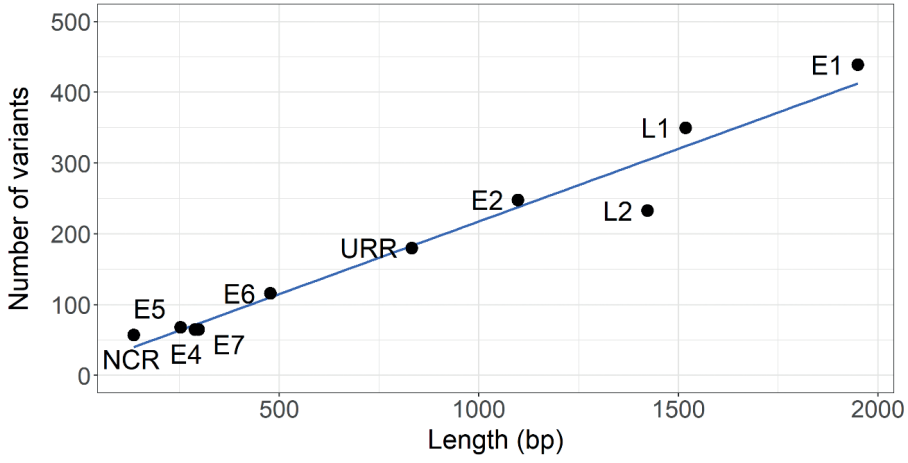


Figure 5: Correlation between the total number of minor nucleotide variants (MNVs) and the length of viral gene regions including URR and NCR. Since certain genes overlap, MNVs can be counted more than once.

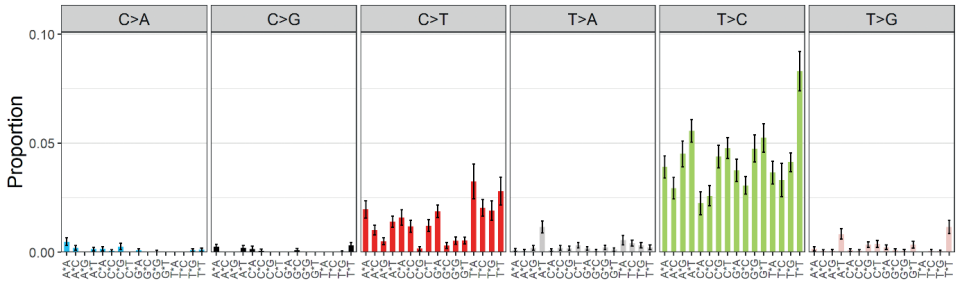


Figure 6: Overall mutational signatures of 1717 minor nucleotide variants (MNVs) in 36 samples. Mutations are classified into six base substitutions and further into 96 trinucleotide substitution types. Mean proportion of each 96 mutational signature was calculated in the samples. Error bars represent the standard error of the mean.

Discussion

Using the highly sensitive TaME-seq assay, we investigated consecutive HPV16 positive samples from the same infection. Our data suggests the presence of numerous HPV16 MNVs. Consensus sequences (major nucleotide variants) were conserved over time (up to two years follow-up), in line with previous results from this cohort [10]. The detection of MNVs correlated with depth of coverage, which in turn and as to be expected, correlated strongly with sample viral load. The distribution of synonymous and non-synonymous MNVs across the genome appeared uniform and therefore gave no grounds for interpreting selection. Furthermore, MNVs are generally greatly outnumbered by the consensus type, which would be available for transcription of functional proteins. At this MNV level we cannot therefore interpret any substitution rates. Further studies using samples with lesions of varying degrees are required to study the dynamics of and associations between specific MNVs and carcinogenesis.

From 15 HPV16 positive infections (36 samples), we identified a total of 1717 polymorphic positions. Per infection we found on average 48 MNVs/genome (range 15–82) using the 1% frequency cutoff. Our study coincides by magnitude with findings reported by de Oliveira *et al.* (5–125 MNVs/genome, 1% cutoff) [15], as well as a study investigating HPV16/52/58 MNVs in CIN1+ by Hirose *et al.* (0–85 MNVs/genome, 0.5% cutoff) [38, 39]. Hirose and colleagues further found that the number of HPV16 variants negatively correlated with histological grade. On average, we observe more variants than Hirose *et al.*, which may be in part due to methodological differences, but likely also due to the age group from which our samples were obtained.

Although the number of MNVs identified is comparable to other studies, the nature of the mutation profiles differs. We find that the overall majority (67%) of MNVs were T>C changes, whereas other studies point to a higher frequency of C>T mutations, which we find the second-most abundant (23%). The ratios of T>C mutations against all MNVs are very consistent in our data irrespective of how they were called. Although TaMe-seq is designed with high primer

density to cover the entire HPV16 genome (52 in total), it is not designed to completely cover the genome with both the two F and R reactions separately. Despite this, nearly half (n = 809/1717) of the called MNVs are found in overlapping regions obtained from the two reactions independently. Most of these (n = 625) are called in either one of the F or R reactions suggesting that they are either below the 1% frequency cutoff in the other reaction, stochastically amplified from a variant pool by only one of the reactions, or noise. The T > C mutation ratio is lowest in these unpaired MNVs (63%) and highest (73%) in those that are called by both the F and R reactions (11% of total MNVs, n = 184/1717). This is the opposite of what could be expected if T > C mutations were erroneously called, assuming that MNVs independently detected by the F and R reactions confirm each other. The probability of falsely calling the same MNV in two separate reactions is extremely small. Therefore, the derived mutation ratios support the overall finding that T > C mutations dominate the MNVs in our samples. The origin of these T > C mutations remains to be explored, particularly with a focus on early infection events and influence from genome dynamics, DNA repair and viral replication.

The C > T mutation profile is associated with APOBEC activity [12, 38]. Over time, APOBEC-related C > T changes accumulate in progressing infections, resulting in mutation patterns observed in CIN1 + materials [12, 38]. Our findings imply that APOBEC activity could manifest at later developmental stages of infections than those included in this study. Interestingly, in our dataset we find one infection that shifts over time from a T > C heavy mutation profile, to a C > T heavy mutation profile (444086, Figure S2), suggesting APOBEC activity. This is further supported by the observation that the C > T mutations in the last collected sample of this infection are almost exclusively in the 5´-TC dinucleotide context which is the preferred APOBEC3A and APOBEC3B motif [40]. MNVs were generally not recaptured in consecutive samples. This may be due to sampling of random fractions of the low frequency MNV for each sample and potentially changes in HPV genome dynamics over time. Despite this, 35 MNVs could be detected repeatedly in follow-up samples. Although the numbers are small, it is noticeable that the T > C ratio is lower (44%) and the C > T ratio higher (34%) in these persistent MNVs relative to the overall distribution of mutations (67 and 28%, respectively). Although this dataset is too small to make firm statements, it is tempting to speculate that an APOBEC footprint accumulates, and therefore becomes more easily detectable in the viral pool over time. This does not necessarily occur from selection but from persisting APOBEC activity. In this study, we repeatedly identified (1–7 times) 45 MNVs at 21 polymorphic sites. These sites overlap with a subset of HPV16 SNPs reported by Mirabello *et al.*, which are significantly associated with disease outcome at the population level [12]. Here, they are identified at the minority level. Although, the biological relevance of low frequency variants is yet to be determined, changes in MNV frequency over time might be an indication of microevolution linking to disease progression. This study presents a first look at the development of MNVs over time. Since previous knowledge on this subject is scarce, a number of unknowns become apparent. Currently, we do not fully comprehend the origin or interplay of minority variants. Variants with similar fitness could be originating naturally over time within hosts, who could then transmit them during intercourse. The role of repeated exposure is also unknown and could lead to an increase of variant diversity for each exposure event. Importantly, the detection of abundant intra host MNVs does not challenge the well-established slow evolution of HPVs but

rather increases our understanding of the variable HPV16 genome substrate that can be available for natural selection and evolution at the population level. Future research is required to unravel the fundamentals of HPV variant genesis and their role in transmission and establishment of new infections.

One of the strengths of this study is the use of TaME-seq for deep whole HPV genome sequencing. A comparison of consensus sequences obtained using TaME-seq with previously described Sanger sequencing data showed similar results [10]. In addition, the robustness and reliability of the bioinformatics pipeline, calling mutation profiles from raw sequence data, was controlled by reanalysis of the data from Hirose *et al.* [38], producing excellent compatibility. Finally, our method enabled us to detect 11% of the called MNVs independently in overlapping reads obtained from the two amplification reactions (F and R). Using these, we compared mutational profiles to the whole dataset and similar distributions of mutations were observed.

The design and method used in this study carry some limitations. TaME-seq genome coverage varied between samples and strongly correlated with the initial HPV load. Since overlapping high-resolution data is required to compare MNVs at different time points of an infection, sample inclusion was limited to ≥100× coverage across >45% of the genome. Consequentially, the mutational patterns observed in this study are often observed on stretches of DNA rather than whole-genome results. It is worth noting that the mutational profiles described in the present analysis, reflect the complete population of HPV16 variants in each sample. No distinction is made between potential co-infections of the same type to prevent potential bias. To compensate for varying viral load of the input material on the resulting sequencing coverage, a downsampling analysis was performed of high-coverage samples, which showed similar results to the original analysis. Therefore, we expect sequence coverage differences to be of limited influence on the observed mutation patterns. However, one 200 nt genomic region was poorly covered in all samples (position 4800–5000), possibly due to scarcity of aligning primers. Potential MNVs in this region may therefore be underreported. One sample with the highest coverage (>10 fold higher than most other samples) and viral load, reported the least number of MNVs (n=15). This illustrates how MNVs may not reach 1% frequency against a massive backdrop of major variants in a competitive amplification step.

MNVs were generally found to differ between consecutive samples. The identification of a number of MNVs which were conserved in consecutive samples (Table S3) suggests that this is at least partially caused by sequence coverage and depth. Uncommon MNVs around the detection cutoff will vary in detection and frequency due to PCR and sequencing stochasticity. In addition, the sequencing resolution dictates the number of variants detected from an expected larger mutational pool. It is likely that highly prevalent MNVs are more frequently detected than MNVs around the detection cutoff, although a correlation between MNV prevalence and consecutive detection could not be confirmed for our dataset. It is likely that each sample preparation step leads to a selection of MNVs from the total pool, making redetection of MNVs over time difficult. Furthermore, biological differences between baseline and follow-up samples account for a large portion of MNVs that could not be repeatedly detected. A high viral load at baseline suggests that many MNVs can be detected, while a low viral load at follow-up suggests that only a limited number could be detected. This could explain how often even prevalent MNVs

could not be detected in follow-up samples. To our knowledge, this dataset is among the first to describe MNVs in follow-up samples, implying that there could be methodological inefficiencies in the redetection of MNVs from follow-up samples. Further research is required to determine the optimal approach for this.

In this study, QIAGEN Multiplex PCR kit with HotStar Taq DNA Polymerase was used, which, like other polymerases lacking proofreading, could introduce a T > C prone error bias [41]. Additionally, some of the observed transitions could be caused by the Illumina platform. However, as described in the methods section, the use of paired-end reads and a cutoff for calling minor variants (> 1%) should minimize bias from these sources. Furthermore, MNVs in 35 individual genome positions were detected repeatedly in consecutive samples and 138 MNVs in both the separate F and R amplification reactions, suggesting robustness for our observations.

The samples used here were obtained from a retrospective cohort study, which was initially aimed at identifying *Chlamydia trachomatis* infections, and later adapted for HPV purposes [17]. Due to the age of the women recruited for this study (16–29 years old), and the fact that they were recruited for *C. trachomatis* purposes, it is unlikely that the study participants have high-grade cytological malignancies, although this could not be confirmed. The longitudinal nature of this study combined with our inclusion criteria, also means that sample size is relatively small. Since this study was originally conducted to assess *C. trachomatis* status, an effect of such infections might be apparent in the mutation rates of the samples tested in the present analysis. However, since only one of the fifteen infections analyzed here was *C. trachomatis* positive, we could not compare mutation rates between *C. trachomatis* positive and negative individuals.

In summary, this study reports a multitude of MNVs observed through whole genome, deep sequencing of HPV16 infection with longitudinal follow-up. The mutation profiles identified in this study suggested non-APOBEC-related pathways causing mutations in HPV16 infections in young women. Most MNVs were detected incidentally, however, some MNVs could be detected separately or repeatedly over time, suggesting robustness in mutational profiles and at least partial conservation of MNVs. Some of the MNVs identified repeatedly were associated with malignant infection outcomes in other studies, potentially suggesting clinical relevance in longitudinal tracking of MNVs.

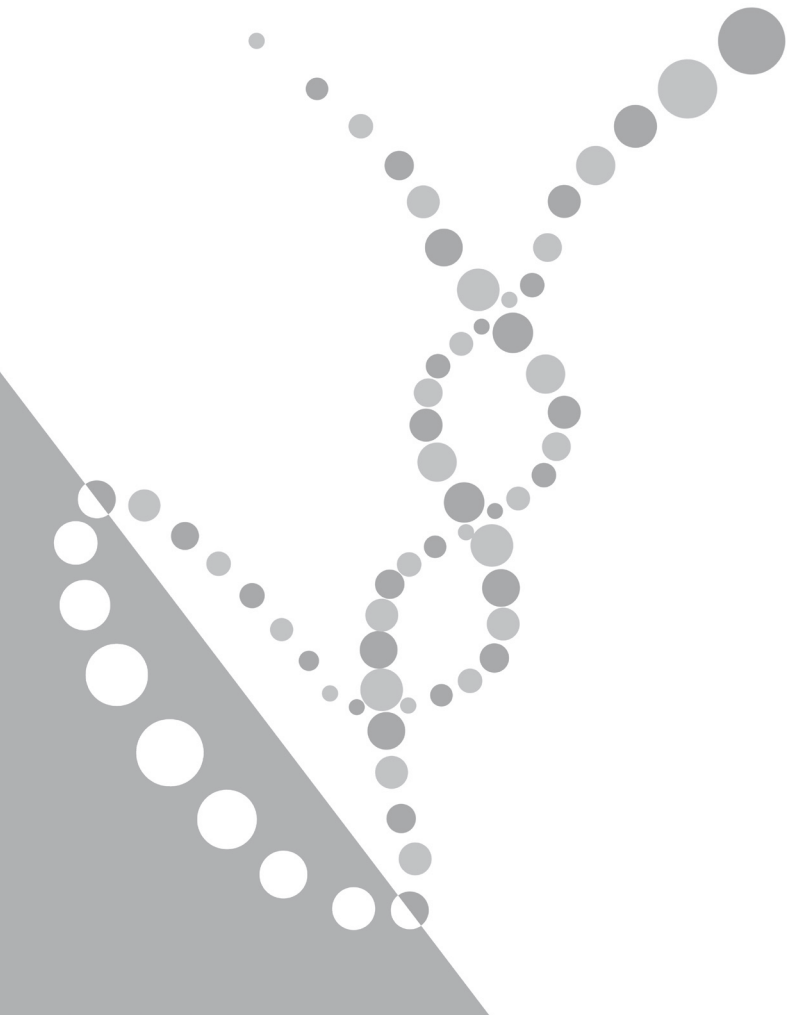
References

1. Baseman, J.G. and L.A. Koutsky, *The epidemiology of human papillomavirus infections.* J Clin Virol, 2005. **32 Suppl 1**: p. S16–24.
2. Walboomers, J.M., et al., *Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.* J Pathol, 1999. **189**(1): p. 12–9.
3. Rositch, A.F., et al., *Patterns of persistent genital human papillomavirus infection among women worldwide: a literature review and meta-analysis.* Int J Cancer, 2013. **133**(6): p. 1271–85.
4. Van Doorslaer, K., *Evolution of the papillomaviridae.* Virology, 2013. **445**(1–2): p. 11–20.
5. Bravo, I.G. and M. Felez-Sanchez, *Papillomaviruses: Viral evolution, cancer and evolutionary medicine.* Evol Med Public Health, 2015. **2015**(1): p. 32–51.
6. Burk, R.D., A. Harari, and Z. Chen, *Human papillomavirus genome variants.* Virology, 2013. **445**(1–2): p. 232–43.
7. Chen, A.A., et al., *Human Papillomavirus 18 Genetic Variation and Cervical Cancer Risk Worldwide.* J Virol, 2015. **89**(20): p. 10680–7.
8. Mirabello, L., et al., *HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women.* J Natl Cancer Inst, 2016. **108**(9).
9. Cullen, M., et al., *Deep sequencing of HPV16 genomes: A new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection.* Papillomavirus Research, 2015.
10. van der Weele, P., C. Meijer, and A.J. King, *Whole-Genome Sequencing and Variant Analysis of Human Papillomavirus 16 Infections.* J Virol, 2017. **91**(19).
11. van der Weele, P., C. Meijer, and A.J. King, *High Whole-Genome Sequence Diversity of Human Papillomavirus Type 18 Isolates.* Viruses, 2018. **10**(2): p. 68.
12. Mirabello, L., et al., *HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis.* Cell, 2017. **170**(6): p. 1164–1174 e6.
13. Dube Mandishora, R.S., et al., *Genotypic diversity of anogenital human papillomavirus in women attending cervical cancer screening in Harare, Zimbabwe.* J Med Virol, 2017. **89**(9): p. 1671–1677.
14. Dube Mandishora, R.S., et al., *Intra-host sequence variability in human papillomavirus.* Papillomavirus Res, 2018. **5**: p. 180–191.
15. de Oliveira, C.M., et al., *High-level of viral genomic diversity in cervical cancers: A Brazilian study on human papillomavirus type 16.* Infect Genet Evol, 2015. **34**: p. 44–51.
16. Doorbar, J., et al., *The biology and life-cycle of human papillomaviruses.* Vaccine, 2012. **30 Suppl 5**: p. F55–70.
17. Kang, S.D., et al., *Effect of Productive Human Papillomavirus 16 Infection on Global Gene Expression in Cervical Epithelium.* J Virol, 2018. **92**(20).
18. Chapman, J.R., M.R. Taylor, and S.J. Boulton, *Playing the end game: DNA double-strand break repair pathway choice.* Mol Cell, 2012. **47**(4): p. 497–510.
19. Moody, C., *Mechanisms by which HPV Induces a Replication Competent Environment in Differentiating Keratinocytes.* Viruses, 2017. **9**(9).
20. Sanjuan, R. and P. Domingo-Calap, *Mechanisms of viral mutation.* Cell Mol Life Sci, 2016. **73**(23): p. 4433–4448.

21. Koito, A. and T. Ikeda, *Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases.* Front Microbiol, 2013. **4**: p. 28.
22. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415–21.
23. Wallace, N.A. and K. Munger, *The curious case of APOBEC3 activation by cancer-associated human papillomaviruses.* PLoS Pathog, 2018. **14**(1): p. e1006717.
24. Warren, C.J., et al., *APOBEC3A functions as a restriction factor of human papillomavirus.* J Virol, 2015. **89**(1): p. 688–702.
25. Mori, S., et al., *Human Papillomavirus 16 E6 Upregulates APOBEC3B via the TEAD Transcription Factor.* J Virol, 2017. **91**(6).
26. Burns, M.B., et al., *APOBEC3B is an enzymatic source of mutation in breast cancer.* Nature, 2013. **494**(7437): p. 366–70.
27. Burns, M.B., N.A. Temiz, and R.S. Harris, *Evidence for APOBEC3B mutagenesis in multiple human cancers.* Nat Genet, 2013. **45**(9): p. 977–83.
28. Lagström, S., et al., *TaME-seq: An efficient sequencing approach for characterisation of HPV genomic variability and chromosomal integration.* Scientific Reports, 2019. **9**(1).
29. Mollers, M., et al., *Prevalence, incidence and persistence of genital HPV infections in a large cohort of sexually active young women in the Netherlands.* Vaccine, 2013. **31**(2): p. 394–401.
30. van den Broek, I.V., et al., *Systematic selection of screening participants by risk score in a Chlamydia screening programme is feasible and effective.* Sex Transm Infect, 2012. **88**(3): p. 205–11.
31. van den Broek, I.V., et al., *Evaluation design of a systematic, selective, internet-based, Chlamydia screening implementation in the Netherlands, 2008–2010: implications of first results for the analysis.* BMC Infect Dis, 2010. **10**: p. 89.
32. Kleter, B., et al., *Development and clinical evaluation of a highly sensitive PCR-reverse hybridization line probe assay for detection and identification of anogenital human papillomavirus.* J Clin Microbiol, 1999. **37**(8): p. 2508–17.
33. Kleter, B., et al., *Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses.* Am J Pathol, 1998. **153**(6): p. 1731–9.
34. van der Weele, P., et al., *Correlation between viral load, multiplicity of infection, and persistence of HPV16 and HPV18 infection in a Dutch cohort of young women.* J Clin Virol, 2016. **83**: p. 6–11.
35. Kozich, J.J., et al., *Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform.* Appl Environ Microbiol, 2013. **79**(17): p. 5112–20.
36. Van Doorslaer, K., et al., *The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis.* Nucleic Acids Res, 2013. **41**(Database issue): p. D571–8.
37. Kim, D., B. Langmead, and S.L. Salzberg, *HISAT: a fast spliced aligner with low memory requirements.* Nat Methods, 2015. **12**(4): p. 357–60.
38. Hirose, Y., et al., *Within-Host Variations of Human Papillomavirus Reveal APOBEC Signature Mutagenesis in the Viral Genome.* J Virol, 2018. **92**(12).
39. Kukimoto, I., et al., *Genetic variation of human papillomavirus type 16 in individual clinical specimens revealed by deep sequencing.* PLoS One, 2013. **8**(11): p. e80583.

40. Warren, C.J., et al., *Roles of APOBEC3A and APOBEC3B in Human Papillomavirus Infection and Disease Progression.* Viruses, 2017. **9**(8).
41. Brandariz-Fontes, C., et al., *Effect of the enzyme and PCR conditions on the quality of high-throughput DNA sequencing results.* Sci Rep, 2015. **5**: p. 8056.

Chapter 7



HPV16 variant analysis in primary and recurrent CIN2/3 lesions demonstrates presence of the same consensus variant

Pascal van der Weele[1,2], Audrey J. King[1], Chris J.L.M. Meijer[2], Renske D.M. Steenbergen[2]

[1]National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Antonie van Leeuwenhoeklaan 9, Bilthoven, the Netherlands
[2]Amsterdam UMC, Vrije Universiteit, Department of Pathology, Cancer Center Amsterdam, De Boelelaan 1117, Amsterdam, the Netherlands

Papillomavirus Res. 2019 Jun; 7: 168–172

Abstract

Introduction Recurrent cervical intraepithelial lesions (rCIN2/3) after treatment of CIN2/3 occur in 5–15% of cases. rCIN2/3 can result from incomplete resection of CIN2/3, where the same HPV type and variant remains present. rCIN2/3 could also occur following a new infection with a different HPV variant of the same HPV type as the initial lesion. This study investigates HPV16 consensus variants in paired HPV16 positive scrapes from baseline CIN2/3 and rCIN2/3 lesions.

Methods Paired HPV16 positive cervical scrapes of women with CIN2/3 at baseline and rCIN2/3 6 or 12 months after treatment were selected for whole-genome amplification and Illumina sequencing. Sequences were compared and nucleotide changes over time were characterized.

Results From 14 paired samples, 10 had identical consensus variants in baseline CIN2/3 and rCIN2/3. Four paired samples showed one to three nucleotide variations at recurrent disease compared to baseline.

Conclusion Identical or nearly identical HPV16 consensus variants were found in scrapes of paired HPV16 positive baseline CIN2/3 and rCIN2/3 lesions after treatment, suggesting no need for HPV variant analysis when the same HPV type is found in both lesions. These results argue for either incomplete excision of baseline CIN2/3 or inability of clearance of the original HPV infection.

Keywords Whole-genome sequencing; HPV16; HPV genome variants; CIN; rCIN; recurrent infection

Introduction

Persistent infection with an oncogenic human papillomavirus (HPV) type is an essential requirement for the development of cervical cancer [1]. However, most HPV infections are cleared by the immune system within one to two years after initial infection [2]. Of the currently known oncogenic HPV types, HPV16 and HPV18 cause around 70% of all cervical cancer cases worldwide [3].

The progression from initial HPV infection to cervical cancer occurs via precursor lesions (cervical intraepithelial neoplasia, CIN grade 1 to 3) and may take decades [4, 5]. Treatment of CIN2/3 lesions is performed by ablative or excisional treatment. Women treated for CIN2/3 lesions have a 5–15% risk of developing recurrent high-grade lesions (rCIN2/3) within two years post-treatment [6, 7]. Consequently, women undergoing treatment for CIN2/3 are closely monitored in the post-treatment period, before they return to a regular screening routine [7, 8]. In the Netherlands women are tested by cytology and HPV co-testing at 6, 12 and 24 months post-treatment [9]. The 12 months visit can be omitted when the 6-month visit shows absence of HPV and normal cytology. After three consecutive negative co-tests, the woman is referred back to screening programme [7, 9].

rCIN2/3 represents a heterogeneous group of lesions, consisting of either residual CIN2/3 or a new CIN2/3 lesion. Residual CIN2/3 is a possible consequence of incomplete excision of the original CIN2/3 lesion, characterized by the same HPV type in the rCIN2/3 as in the baseline lesion. A new CIN2/3 lesion would occur from a newly acquired HPV infection by a different type (type switch), or an infection with a different variant of the same HPV type present in the baseline lesion [10]. Post-treatment surveillance should ideally differentiate residual from incident lesions, as women with residual lesions are in need of immediate treatment. Women with incident lesions may benefit from a more conservative approach due to a lower cancer risk [11].

In a recent multicenter post-treatment surveillance study, most baseline CIN2/3, rCIN3 and a subset of rCIN2 harbored HPV16 by genotyping [11]. From a clinical perspective, this poses the question whether the rCIN2/3 was caused by a newly acquired HPV16 infection, or by the same infection detected at baseline, which resulted in CIN2/3. Here, we employ whole genome next-generation sequencing of HPV16 to identify and compare consensus variants in paired baseline CIN2/3 and rCIN2/3 cervical scrapes.

Table 1: Characteristics of patients included in this study.

Study number	Histology at 6 or 12 months post treatment	Months post-disease diagnosis	Nucleotide changes
10	CIN2	7	3723 T>G
203	CIN2	5	-
402	CIN3	14	-
410	CIN3	11	-
422	CIN3	12	3800 C>G, 7702 G>C
638	CIN3	6	-
648	CIN3	6	-
665	CIN3	7	-
669	CIN2	5	3800 C>G
672	CIN2	12	-
680	CIN3	6	-
681	CIN2	6	-
834	CIN3	8	-
872	CIN2	12	3375 C>T, 3575 C>T, 5306 C>T

Analysis was performed on HPV16 positive cervical scrapes for both baseline CIN2/3 and rCIN2/3.

Materials and Methods

Sample selection

Cervical scrapes from women (aged 18+ years) with CIN2/3 derived from a multicenter study (SIMONATH) which has been described earlier [11, 12] and who were scheduled for LLETZ (Large Loop Excision of the Transformation Zone) treatment of CIN2/3 were included. Baseline scrapes were obtained between two weeks before up to immediately before LLETZ. In addition, preceding scrapes from women with a CIN2/3 found in the LLETZ material, without an additional biopsy could also be included. An additional scrape was taken from patients prior to treatment. For this study, only baseline CIN2/3 and rCIN2/3 scrapes were tested. Women treated for HPV16 positive baseline CIN2/3, and with HPV16 positive rCIN2/3 at six or 12-month follow-up were selected. Based on sample availability, a total of 14 HPV16 positive scrape pairs were tested, corresponding to 14 baseline CIN2/3 and 14 rCIN2/3 (6 rCIN2, 8 rCIN3), as shown in Table 1. The SIMONATH study was approved by the Medical Ethical Committee (METC) of VUmc (2009/285) and was registered in the Dutch Trial Registry (NTR1964).

HPV DNA detection

HPV detection and genotyping have been described previously for this study [11, 12]. In short, total DNA was isolated from cervical scrapes using the Microlab start platform (Hamilton Robotics, Switzerland) with magnetic beads (Macherey-Nagel, Germany) according to the manufacturer's protocol. HPV DNA was amplified using the GP5+/6+ PCR, followed by detection via enzyme immunoassay readout and genotyping via an in-house reverse line blot [13].

HPV16 whole genome amplification

Complete HPV16 genomes were amplified in ten fragments from selected samples using primers displayed in Table 2. PCR's were performed using AmpliTaq Gold (Thermo Fisher Scientific, United States) according to the manufacturer's protocol. Cycling conditions consisted of an initial incubation of 15 minutes at 95°, followed by 38 cycles of alternating 95° for 15 seconds, 55° for 30 seconds and 72° for 90 seconds, followed by a final elongation step at 72° for 10 minutes. Amplicon integrity was checked using the Lonza FlashGel (Lonza, Switzerland) system. If no product was observed, it was assumed the PCR performed at too little efficiency to observe on gel. Sample dsDNA concentrations were quantitated using the Quant-iT PicoGreen dsDNA assay kit (Thermo Fisher Scientific, United States) according to the manufacturer's protocol to facilitate equimolar pooling of PCR products.

Illumina NGS analysis

Samples were submitted to BaseClear for Illumina HiSeq PE125 sequencing. Raw sequencing data was subjected to trimmomatic 0.36 for quality and adapter trimming (SLIDINGWINDOW:5:25; MINLEN:35), followed by FastQC 0.11.6 and MultiQC 1.3 for quality checking. Trimmed sequences were assembled to a HPV16 (K02718) reference genome using Bowtie2 2.3.4. Assembly files (.sam) were converted to bam and indexed using samtools 1.6. Consensus sequences were extracted from bam files using samtools and seqtk 1.2. Variant calling files (vcf) were generated using Lofreq to assess the presence of variants at heterogeneous positions. The cutoff for minority variants was set at 0.5%. Finally, bed files were extracted using bedtools 2.27.1 to facilitate the generation of coverage plots in R.

Alignment and phylogeny

Consensus sequences were aligned using MUSCLE 3.8.1551. Maximum likelihood phylogenetic inference was performed using IQ-tree 1.5.5. The model finder option (-m MF) was used to identify the best fit model for this study, resulting in the HKY+F+I model. The final alignment was bootstrapped using IQ-tree's ultrafast bootstrapping option (-bb 1000). Alignments were visualized using FigTree 1.4.3.

Results

Sample selection, amplification and sequencing

For this study, fourteen women treated for HPV16 positive CIN2/3, and with HPV16 positive rCIN2/3 at follow-up were selected (Table 1). Amplification of the whole HPV viral genome via overlapping PCR fragments was successful in all cervical scrapes. Subsequent ultra-deep sequencing of the cervical scrapes resulted in very high genome-wide coverage (Figure S1), with a pooled average coverage per genome position of 112.287 (median: 104.130, minimum: 2045, maximum: 250.000), allowing for reliable assessment of infection variants.

Sequence comparison and characterization of nucleotide changes

Comparison of ultra-deep sequencing results of all fourteen paired samples showed (near) identical consensus variants at baseline and at recurrent disease, as shown in a maximum likelihood plot of the consensus sequence data in Figure 1. Out of fourteen infections included in this study, ten are identical at the consensus level for both baseline and recurrent disease. In three patients, a single nucleotide polymorphism (SNP) was found at recurrent disease, at nucleotides 3723 and twice at 3800, respectively. In the remaining patient, three SNPs were detected at recurrent disease, at nucleotides 3375, 3575 and 5306. The sequence variations are listed in Table 1. A comparison of variant sequence, variant counts at each variable position and trinucleotide context, is presented in Table 3.

Discussion

In this study, we have investigated HPV16 variants in paired cervical scrapes of baseline CIN2/3 and rCIN2/3 at six or 12-month follow-up. Consensus variant analysis suggested that baseline CIN2/3 and rCIN2/3 are (near) identical in all cases. From a clinical perspective, this implies that the infection causing the initial CIN2/3 lesion was either not completely removed during treatment and resulted in a recurrent lesion, or was caused by a novel infection with a (nearly) identical HPV16 variant. Our findings imply that in clinical practice, conventional PCR and genotyping is sufficient to detect type switches, since in the case of rCIN with the same HPV type as baseline CIN, both are likely caused by the same HPV variant.

Our comparison of consensus sequences showed that 10 out of 14 patients have identical HPV16 consensus genomes at baseline CIN2/3 and rCIN2/3. For these patients, the most parsimonious explanation is that the infection causing the initial CIN2/3 lesion also caused the rCIN2/3 lesion. The remaining four patients had consensus sequences that were nearly identical (1 and 3 nucleotides difference).

Previous studies have suggested that whole genome sequences differing ≥2 nucleotides could be considered unique variants, while sequences differing <2 nucleotides cannot be reliably discriminated from each other [14]. In two patients, the rCIN2/3 positive HPV16 infection

showed a one nucleotide difference compared to the HPV16 infection identified at baseline CIN2/3. Two more infections were found in which CIN2/3 and rCIN2/3 differed by two or three nucleotides. Considering the conservation of the HPV genome over time and the diversity of variants circulating in populations, the variants in the rCIN2/3 lesions most likely originated from the variants causing the initial CIN2/3. By definition, some, or all of these infections could be reinfections with unique variants, however, considering the plethora of different HPV16 variants circulating in populations [14, 15] it is improbable that one would be repeatedly infected with so closely related variants.

Table 2: Primer sequences used in this study with references of origin.

Forward	Sequence 5'-3'	Reference	Reverse	Sequence 5'-3'	Reference
F7869	GGTTACACATTTACAAGCAAC	*	R1312	ACATGGTGTTTCAGTCTCATGGC	[14]
F6835	CTGTGCAAAATAACCTTAACTGC	[14]	R162	GCAGCTCTGTGCATAACTGTG	[14]
F5492	TATAACTGACCAAGCTCCTTCA	[15]	R6599	TTATTGTGGCCCTGTGC	[16]
F6201	GAACACTGGGGCAAAGGATC	[17]	R6890	GAATTCATAGAATGTATGTATGTC	[14]
F3701	CGTCTACATGGCATTGGAC	[18]	R5024	AAGCAGGGTCTACAACTTTAAC	[14]
F4930	AACTAGTAGCACACCCATACCA	[14]	R5725	CGTGCAACATATTCATCCGT	[17]
F2529	CAATTTAAGAAATGCATTGGATGG	[14]	R3551	GTCTGGCTCTGATCTTGGTC	[14]
F3387	GTCAGGTAATATTATGTCCTACA	[14]	R4321	TGCAGAACGTTTGTGTCGCATT	[14]
F901	ACGGGATGTAATGGATGG	*	R1780	ATCATACACATTGGAGACACA	[18]
F1832	CAATGTGTATGATGATAGAGCC	[14]	R2915	AATAGTCTATATGGTCACGTAGG	[14]

Primers developed for this study are marked with *.

Assessment of variants at the variable positions showed that in five out six cases, the consensus nucleotide at rCIN was already present as a minority at baseline CIN. In addition, five out of six SNPs are possible apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) mutations (C > G and C > T in a tCn trinucleotide motif) [16]. APOBEC mutations are suggested to accumulate naturally as a host response against viruses, and have been shown to occur on the HPV genome [14, 17, 18]. This could be particularly relevant for the patient with three nucleotide changes between CIN2/3 and rCIN2/3 (872), as all three changes are possible APOBEC mutations. Combined, these findings suggest that the variable nucleotides originated as mutation events, but the effect of treatment on their prevalence cannot be assessed, since this study lacks the appropriate controls and power for such an analysis.

Overall, most variations occur in the *E2* gene. The SNP at position 3375 is in the *E2* hinge region, which is hypervariable between HPV types, and is not clearly associated with any function [19]. Position 3575 has been described as an integration site of HPV16 [20], although it is unknown how a nucleotide shift at this position affects this. Variations at positions 3723 and 3800 lead to changes in the DNA binding domain of *E2* [19], and 3800 C > G has been identified in a cervical cancer case in India [21]. In *L2* 5306 C > T was found, although this position does not encode any

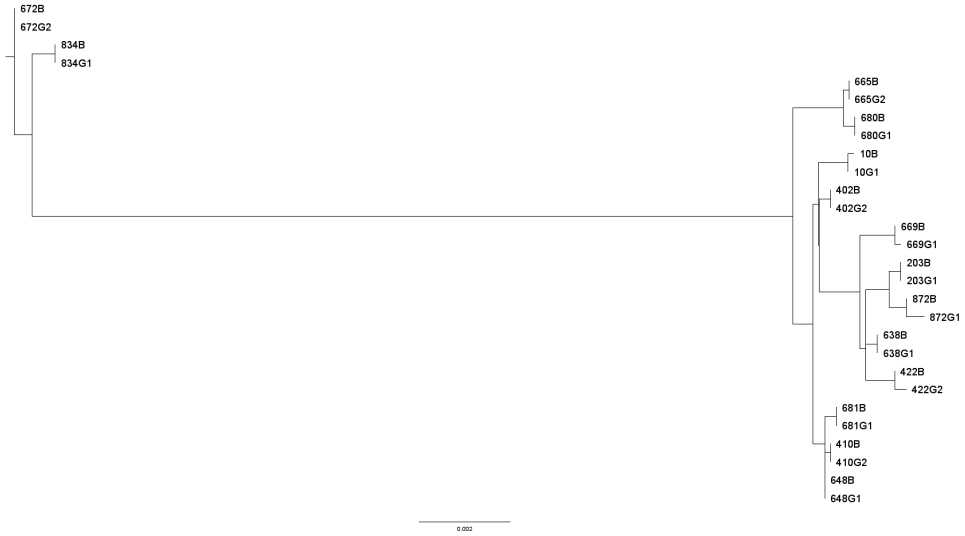


Figure 1: Maximum likelihood tree of consensus sequences obtained in this study. Baseline (B) and follow-up (G1/2) samples are shown to cluster close or identical to each other. X-axis shows genetic distance between samples.

Table 3: Nucleotide counts at variable positions between baseline CIN and rCIN. In addition, the trinucleotide context in which each variation occurs is given.

Sample	Nucleotide position	Total coverage	Major nucleotide	Coverage	Minor nucleotide	Coverage	Trinucleotide context
10B	3723	193,552	T	169,840	G	23,467	gTa
10G1	3723	167,524	G	167,242	-	-	gGa
422B	3800	163,308	C	146,166	G	16,826	tCa
422G2	3800	153,655	G	146,319	C	7,123	tGa
669B	3800	102,591	C	102,279	-	-	tCa
669G1	3800	138,622	G	131,277	C	7,121	tGa
872B	3375	99,262	C	56,321	T	42,874	tCc
872G1	3375	105,095	T	93,927	C	11,124	tTc
872B	3576	92,831	C	50,678	T	42,041	tCa
872G1	3576	106,969	T	96,364	C	10,557	tTa
872B	5306	111,061	C	59,651	T	50,363	tCa
872G1	5306	118,822	T	98,347	C	20,445	tTa

known epitopes [22]. URR position 7702 does not encode for any known promoter sites [23]. Currently it is unclear what the exact role is of any of the nucleotide changes observed between baseline CIN2/3 and rCIN2/3, but the fact that most are found in *E2*, which is associated with *E6/E7* regulation [19] is striking and warrants close monitoring of these infections.

While the method employed in this study generates ultra-high resolution sequencing data across most of the HPV16 genome, upon observing Figure S1, it becomes apparent that there are some coverage dips around nucleotide positions 1800, 4100, 7000 and 7800. This suggests sub-optimal primer design, although coverage does not drop below 2000 on average. Further optimization could yield a more equal genome wide result, although this is not required for this study. For one infection (669), coverage dropped to 1–10× across a 600bp fragment for the rCIN sample, while at these positions (positions 4251–4871) a 100× coverage was obtained for the baseline sample. This could cause discrepancies between baseline CIN2/3 and rCIN2/3. However, as seen in Figure 1, samples from this infection differ by only one nucleotide, at position 3800, where both samples have >100.000 coverage, ruling out that the difference in sequence is caused by low coverage. Potentially, this could be caused by the presence of both episomal and integrated virus in the CIN lesion at baseline. It may be speculated that following resection, the episomal fraction may have been cleared, leaving only the integrated fraction in the recurrent lesion, or vice versa. We can also not exclude that the low coverage is potentially caused by a deletion in the genome of this specific variant. The method employed in this study could potentially detect deletions within amplicons, but only if the deleted area does not overlap with a primer site.

The results from this study focus primarily on the comparison of consensus sequences to identify causative variants for CIN and rCIN. However, there could be a change in the distribution of minority variants. Although a detailed assessment of the effects of treatment on minority variant diversity would be of interest, this is beyond the scope of the present study.

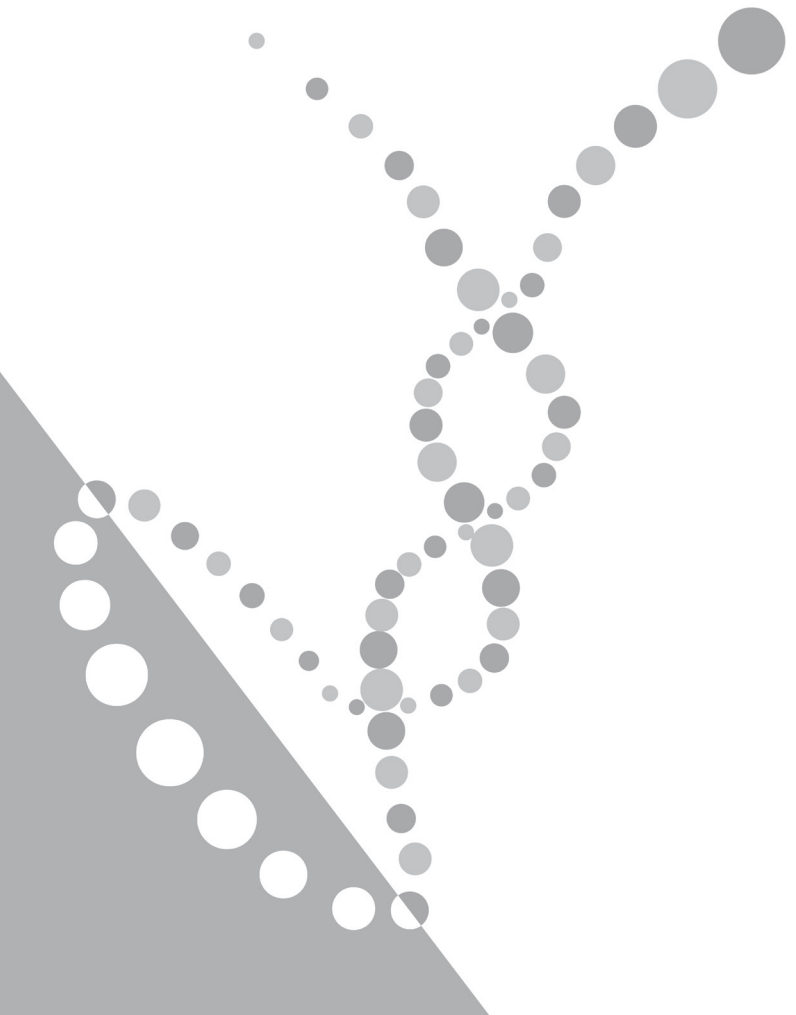
In summary, we showed that when paired scrapes from women with baseline CIN2/3 and post-treatment rCIN2/3 at 6 or 12 months follow-up are both HPV 16 positive, the same variant is responsible for CIN2/3 and rCIN2/3. Consequentially, in clinical practice no need exists for variant analysis when the same HPV type is found at baseline and at recurrent disease. Our findings suggest possible inadequacies in treatment procedures enforced upon detection of CIN2/3 or persistence of the original HPV infection.

References

1. Walboomers, J.M., et al., *Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.* J Pathol, 1999. **189**(1): p. 12–9.
2. Rositch, A.F., et al., *Patterns of persistent genital human papillomavirus infection among women worldwide: a literature review and meta-analysis.* Int J Cancer, 2013. **133**(6): p. 1271–85.
3. de Sanjose, S., et al., *Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study.* Lancet Oncol, 2010. **11**(11): p. 1048–56.
4. Vink, M.A., et al., *Clinical progression of high-grade cervical intraepithelial neoplasia: estimating the time to preclinical cervical cancer from doubly censored national registry data.* Am J Epidemiol, 2013. **178**(7): p. 1161–9.
5. Steenbergen, R.D., et al., *Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions.* Nat Rev Cancer, 2014. **14**(6): p. 395–405.
6. Melnikow, J., et al., *Cervical intraepithelial neoplasia outcomes after treatment: long-term follow-up from the British Columbia Cohort Study.* J Natl Cancer Inst, 2009. **101**(10): p. 721–8.
7. Kocken, M., et al., *Risk of recurrent high-grade cervical intraepithelial neoplasia after successful treatment: a long-term multi-cohort study.* Lancet Oncol, 2011. **12**(5): p. 441–50.
8. Kocken, M., et al., *High-risk human papillomavirus testing versus cytology in predicting post-treatment disease in women treated for high-grade cervical disease: a systematic review and meta-analysis.* Gynecol Oncol, 2012. **125**(2): p. 500–7.
9. Integraal Kankercentrum Nederland. *Richtlijn CIN, AIS EN VAIN*. 2015; Available from: https://www.onco-line.nl/cin-ais-en-vain.
10. Bleeker, M.C., C.J. Meijer, and J. Berkhof, *Follow-up after treatment for cervical intraepithelial neoplasia.* BMJ, 2012. **345**: p. e7186.
11. Uijterwaal, M.H., et al., *Performance of CADM1/MAL-methylation analysis for monitoring of women treated for high-grade CIN.* Gynecol Oncol, 2016. **143**(1): p. 135–142.
12. Polman, N.J., et al., *Good performance of p16/ki-67 dual-stained cytology for surveillance of women treated for high-grade CIN.* Int J Cancer, 2017. **140**(2): p. 423–430.
13. van den Brule, A.J., et al., *GP5+/6+ PCR followed by reverse line blot analysis enables rapid and high-throughput identification of human papillomavirus genotypes.* J Clin Microbiol, 2002. **40**(3): p. 779–87.
14. Mirabello, L., et al., *HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis.* Cell, 2017. **170**(6): p. 1164–1174 e6.
15. van der Weele, P., C. Meijer, and A.J. King, *Whole-Genome Sequencing and Variant Analysis of Human Papillomavirus 16 Infections.* J Virol, 2017. **91**(19).
16. Roberts, S.A., et al., *An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers.* Nat Genet, 2013. **45**(9): p. 970–6.
17. Cancer Genome Atlas Research Network, et al., *Integrated genomic and molecular characterization of cervical cancer.* Nature, 2017. **543**(7645): p. 378–384.
18. Gaglia, M.M. and K. Munger, *More than just oncogenes: mechanisms of tumorigenesis by human viruses.* Curr Opin Virol, 2018. **32**: p. 48–59.
19. McBride, A.A., *The papillomavirus E2 proteins.* Virology, 2013. **445**(1–2): p. 57–79.

20. Liu, Y., et al., *Comprehensive mapping of the human papillomavirus (HPV) DNA integration sites in cervical carcinomas by HPV capture technology.* Oncotarget, 2016. **7**(5): p. 5852–64.
21. Bhattacharjee, B., et al., *Characterization of sequence variations within HPV16 isolates among Indian women: prediction of causal role of rare non-synonymous variations within intact isolates in cervical cancer pathogenesis.* Virology, 2008. **377**(1): p. 143–50.
22. Wang, J.W. and R.B. Roden, *L2, the minor capsid protein of papillomavirus.* Virology, 2013. **445**(1–2): p. 175–86.
23. Lace, M.J., et al., *Upstream regulatory region alterations found in human papillomavirus type 16 (HPV-16) isolates from cervical carcinomas increase transcription, ori function, and HPV immortalization capacity in culture.* J Virol, 2009. **83**(15): p. 7457–66.

Chapter 8



Bivalent Human Papillomavirus (HPV) Vaccine Effectiveness Correlates With Phylogenetic Distance From HPV Vaccine Types 16 and 18

Johannes A. Bogaards[1,2], Pascal van der Weele[1,3], Petra J. Woestenberg[1,4], Birgit H. B. van Benthem[1], and Audrey J. King[1]

[1] Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven
[2] Department of Epidemiology and Biostatistics
[3] Department of Pathology, Vrije Universiteit Amsterdam, Amsterdam University Medical Centres
[4] Care and Public Health Research Institute, Maastricht University Medical Centre, The Netherlands

J Infect Dis. 2019 Jun 4. pii: jiz280

Abstract

To substantiate cross-protection reported across AS04-adjuvanted bivalent human papillomavirus (HPV) vaccine (2vHPV) studies, we reevaluated vaccine effectiveness against type-specific HPV positivity as a function of phylogenetic distance to vaccine target types HPV-16 and -18. We provide evidence of sustained cross-protection up to 8 years postvaccination in a high-risk population in the Netherlands. Moreover, our findings suggest that genomic distance better explains cross-protection than distance measures based on capsid antigens only. Taken together, 2vHPV is predicted to provide partial cross-protection against HPV-31, -33, -35, -45, -52, and possibly -58, that is, acknowledged oncogenic types with close phylogenetic relationships to HPV-16 or -18.

Introduction

The sexually transmitted human papillomavirus (HPV) is considered a necessary factor for development of cervical cancer and is linked to other anogenital and oropharyngeal carcinomas [1]. Papillomaviruses are characterized by genotype, defined as >10% DNA sequence divergence from other known genotypes (generally termed "types") in the *L1* capsid gene [2]. Most HPV-related malignancies are attributable to types 16 and 18. Consequently, first-generation vaccines, based on recombinant expression of L1 in systems yielding virus-like particles (VLPs), focused on HPV-16 and -18, with the quadrivalent vaccine (4vHPV) also containing L1 VLPs of HPV-6 and -11, primarily associated with anogenital warts. As up to 30% of cervical cancer is attributed to oncogenic types other than 16 or 18, achieving broader protection through cross-reactivity or expansion of the range of VLP types is desirable.

Endeavors to expand the range of VLP types have resulted in the second-generation nonavalent vaccine (9vHPV), containing L1 VLPs from those already contained in 4vHPV plus the 5 next most common types in cervical cancer: HPV types 31, 33, 45, 52, and 58. Alternatively, the minor capsid protein L2, though less immunogenic than L1, is potentially an effective target for prophylaxis, as several subdominant protective epitopes of L2 are well conserved between types and broadly cross-protective in animal models. By contrast, the protection elicited by L1 VLPs is generally taken to be type-restricted (ie, reactive with the homologous type) [1].

First-generation HPV vaccines have shown durable type-specific protection for at least a decade [3]. Importantly, this protection is not absolutely type-restricted, because significant cross-protection has been observed against several nonvaccine types, particularly for the AS04-adjuvanted bivalent vaccine (2vHPV) containing L1 VLPs of HPV-16 and -18 only. In the largest phase 3 trial of 2vHPV, cross-protection was described against persistent HPV-6, -31, -33, -45, -51, and -52 infections, and against incident HPV-35 infection. However, findings with regard to nonvaccine types are equivocal, as the 2vHPV trial from Costa Rica reported significant protection against HPV-31, -45, and -52, insignificant protection against HPV-33, and no effect on HPV-51 [4].

Recent population-based studies from the United Kingdom and the Netherlands confirm some cross-protection from 2vHPV in postvaccine surveillance. In Scotland, a decrease in the prevalence of HPV-31, -33, and -45 was observed among women who underwent their first cervical screening within 7 years after initiating a 2vHPV vaccination program [5]. In the Netherlands, significant cross-protection was estimated against HPV-31, -35, -45, and -52 among female visitors to sexually transmitted infection (STI) clinics who reported to be vaccinated, relative to vaccine-eligible controls [6]. Cross-protection from 2vHPV against HPV-6 or -11 has not been replicated in postvaccine surveillance, neither in England [7] nor in the Netherlands [8].

To reconcile the inconsistencies in cross-protection reported across 2vHPV studies and to assess the type-restricted nature of the protection elicited by L1 VLPs, we reevaluated vaccine effectiveness (VE) against type-specific HPV positivity among STI clinic visitors up to 8 years after vaccination as a function of phylogenetic distance to L1 capsid antigens contained in 2vHPV.

Methods

We estimated VE from the Papillomavirus Surveillance Among STI Clinic Youngsters in the Netherlands (PASSYON) study, a biennial cross-sectional survey in the Netherlands, as described before [6], but now with an extra study round and including all genotypes in the SPF10-LiPA25 assay (DDL Diagnostics Laboratory). In brief, women aged 16–24 years, who had been eligible for HPV vaccination since 2009 and visited the STI clinic between 2011 and 2017, provided a vaginal swab that was analyzed using a polymerase chain reaction (PCR)–based assay able to detect 25 HPV types, including 12 acknowledged and 3 possibly oncogenic types (Table 1). We compared type-specific HPV positivity between 1305 self-reported vaccinated (≥1 dose) and 799 unvaccinated women. The self-reported vaccination status was validated by serology among those who also provided blood.

Phylogenetic distance of each genotype to the 2vHPV types used for construction of VLPs was calculated from reference DNA sequences obtained via the papillomavirus episteme, a database of curated papillomavirus genomic sequences [9]. We performed a phylogenetic analysis on L1 amino acid composition using a general Dayhoff matrix for evolutionary change in L1 protein with standard codon model and multiple sequence alignment (www.ebi.ac.uk/Tools/msa/muscle/). In addition, we constructed phylogenetic trees directly from DNA sequences on the basis of *L1* capsid gene, and on the basis of whole-genome sequences (WGS). Unrooted evolutionary trees from L1 protein or DNA sequences were constructed by maximum likelihood with substitution model selection using IQ-TREE version 1.6.0 software (www.iqtree.org/). Phylogenetic distance was calculated from the consensus tree constructed from 1000 ultra-fast bootstrap trees [10]. Finally, we compared dependence on phylogenetic distance to the Hamming distance from aligned L1 sequences, that is, the number of positions at which the corresponding L1 proteins of reference types are different from the amino acids expressed by VLPs in 2vHPV.

We assessed VE as a function of minimum distance to VLP amino acid composition in L1 protein analysis and as a function of minimum distance to HPV-16 or 18 reference sequences in DNA analyses. Because the LiPA25 assay cannot distinguish between types 68, 73, and 97, we omitted these types from statistical analysis. We also omitted HPV-59, as the estimate of cross-protection against this type is potentially hampered by technical issues in the assay [6]. For the remaining types, we fitted a penalized regression spline to the estimates from the logistic mixed model, weighted by the square root of the number of positive test results used in VE estimation (Table 1), as a function of phylogenetic distance. The smoothness of the function was determined by general cross-validation, and confidence intervals (CIs) were obtained through Bayesian approximation. In addition, we performed weighted covariance analyses on the rank values of the various distance measures, stratified by (putative) oncogenicity of HPV types. Statistical analyses were performed with R version 3.5.1.

Table 1. Bivalent Human Papillomavirus (HPV) Vaccine Effectiveness Against Type-Specific HPV Positivity

HPV Type[a]	VE (95% CI)[b]	No., Total (No. Vaccinated/ No. Nonvaccinated)[c]	Reference Genome[d]	Hamming Distance[e]
High-risk				
16	0.92 (.86–.96)	100 (13/87)	gi|333031|lcl|HPV16REF.1	37
18	0.89 (.78–.94)	63 (11/52)	gi|60975|lcl|HPV18REF.1	41
31	0.66 (.51–.77)	129 (50/79)	gi|333048|lcl|HPV31REF.1	102
33	0.41 (.05–.63)	73 (37/36)	gi|333049|lcl|HPV33REF.1	111
35	0.40 (–.03 to .65)	55 (28/27)	gi|396997|lcl|HPV35REF.1	105
39	0.15 (–.19 to .39)	165 (98/67)	gi|333245|lcl|HPV39REF.1	128
45	0.81 (.55–.92)	28 (7/21)	gi|397022|lcl|HPV45REF.1	91
51	−0.24 (–.54 to .01)	522 (345/177)	gi|333087|lcl|HPV51REF.1	180
52	0.36 (.19–.50)	342 (185/157)	gi|397038|lcl|HPV52REF.1	117
56	−0.17 (–.59 to .14)	220 (145/75)	gi|397053|lcl|HPV56REF.1	173
58	0.30 (–.06 to .54)	95 (52/43)	gi|222386|lcl|HPV58REF.1	113
59	−0.95 (−2.17 to −.20)	96 (73/23)	gi|557236|lcl|HPV59REF.1	125
Probable high-risk				
53	0.26 (.05–.43)	332 (189/143)	gi|9627377|lcl|HPV53REF.1	173
66	0.02 (–.26 to .24)	340 (212/128)	gi|1020290|lcl|HPV66REF.1	174
68/73/97	−0.08 (–.63 to .28)	110 (71/39)	gi|71726685|lcl|HPV68REF.1	139
			gi|1491692|lcl|HPV73REF.1	150
			gi|71726694|lcl|HPV97REF.1	89
Low-risk				
6	−0.15 (–.52 to .14)	263 (172/91)	gi|60955|lcl|HPV6REF.1	161
11	−0.07 (−1.00 to .42)	45 (29/16)	gi|333026|lcl|HPV11REF.1	162
34	0.00 (−1.57 to .61)	19 (12/7)	gi|9627334|lcl|HPV34REF.1	154
40	0.08 (–.68 to .50)	46 (28/18)	gi|397014|lcl|HPV40REF.1	163
42	−1.27 (−3.34 to –.19)	57 (45/12)	gi|333211|lcl|HPV42REF.1	160
43	−0.78 (−2.13 to –.01)	67 (50/17)	gi|40804474|lcl|HPV43REF.1	176
44	−0.32 (−1.30 to .24)	61 (42/19)	gi|1020242|lcl|HPV44REF.1	162
54	−0.34 (–.95 to .08)	141 (97/44)	gi|9628437|lcl|HPV54REF.1	143
70	0.02 (−1.03 to .52)	32 (20/12)	gi|1173493|lcl|HPV70REF.1	126
74	0.26 (–.19 to .54)	75 (42/33)	gi|27462483|lcl|HPV74REF.1	168

Abbreviations: CI, confidence interval; HPV, human papillomavirus.

[a]HPV genotypes in the SPF10-LiPA25 assay, with 68 being indistinguishable from 73 and 97.

[b]VE (with 95% CI) was calculated as 1 minus the adjusted odds ratio from a logistic mixed model described in [6].

[c]No. of positive test results (among 1305 vaccinated + 799 nonvaccinated women) used in VE estimation.

[d]Whole-genome reference DNA sequences obtained from the papillomavirus genome database (https://pave.niaid.nih.gov/).

[e]Minimum number of different amino acids between aligned L1 sequences of reference types and virus-like particles in bivalent vaccine.

Results

Type-specific VE estimates are provided in Table 1. The pooled VE against the 2 vaccine types was 91.0% (95% CI, 86.0%–94.2%). Pooled VE against all (possibly) oncogenic types was 25.8% (95% CI, 17.7%–33.2%), whereas pooled VE against nononcogenic types included in the assay was −4.9% (95% CI, −20.7% to 8.8%). We found no indications for dependency of VE on time since vaccination in stratified analyses, comparing women who were offered vaccination < 5 years ago, 5–6 years ago, or 7–8 years ago (Supplementary Table 1).

Overall, there was a clear relationship between VE and phylogenetic distance in L1 protein analysis (Figure 1A). The consistently high cross-protection reported for HPV-45 is due to its close relationship to HPV-18 (Figure 1B). The cross-protection of around 50% against HPV-31 and -35 fits their almost equidistant relationship to HPV-16. Likewise, the cross-protection of around 35% against HPV-33, -52, and -58 is in line with these types having approximately similar phylogenetic distance to HPV-16. Our analysis further supports the notion that the estimate of VE against HPV-59 appears to be an outlier. Of the acknowledged oncogenic types, HPV-51, -56, and -66 are most distantly related to either L1 VLP and least likely to be affected by cross-neutralizing antibodies induced by 2vHPV (Supplementary Figure 1A). The significant rank correlation ($\rho = -0.70$ [95% CI, −.83 to −.51) between VE and phylogenetic distance to L1 VLP in protein analysis was entirely explained by oncogenic types ($\rho = -0.93$ [95% CI, −.95 to −.89]), as no significant correlation was observed for nononcogenic types (Supplementary Figure 1B).

Analysis based on Hamming distance toward L1 VLPs yielded similar results as L1 phylogenetic analysis (Supplementary Figure 2A). Phylogenetic analysis based on the *L1* capsid gene and WGS yielded slightly different depictions but was still comparable to those for L1 protein (Supplementary Figure 2B and 2C). To express the specific association between VE and each phylogenetic distance while controlling for the effect of other measures, we computed their partial rank correlations from the inverse weighted covariance matrix. Apparently, WGS phylogenetic distance to HPV-16 or -18 was the strongest independent determinant of VE ($\rho_{partial} = -0.53$, $P < .01$), with HPV-51, -53, -56, and -66 located around the threshold genomic distance still informative for VE. The nononcogenic types were further distanced from vaccine target types in WGS analysis than in analyses based on capsid antigen only, and the partial rank correlations between VE and L1 distance measures were no longer significant when corrected for genomic distance (Supplementary Table 2).

Discussion

This study provides evidence of sustained cross-protection from 2vHPV up to 8 years postvaccination in a high-risk population. Taken together, 2vHPV is predicted to provide partial cross-protection against HPV types 31, 33, 35, 45, 52, and possibly 58—that is, high-risk types belonging to HPV α-7 (including HPV-18) or α-9 (including HPV-16) species [2]. Of those, HPV-35 and -58 are not frequently reported among the cross-protective types, which may be due to their relative rarity as compared to other cross-protective types. Likewise, VE against HPV-35 and -58 was

not significantly different from zero in this study, and regression analyses based on phylogenetic distance predicted only small to moderate effect size. Of the other (possibly) oncogenic types, cross-protection may extend to HPV-39 but is unlikely for HPV-51, -53, -56, and -66 (ie, high-risk types belonging to HPV α-5 or α-10 species) [2]. Similarly, cross-protection against low-risk types was not observed in our study [8] and is not to be expected on the basis of phylogenetic analyses.

Although there has been concern about the durability of cross-reactivity, so far there is no evidence for the waning of cross-protection from 2vHPV in women who have been vaccinated 3 times, as per initial recommendation [5, 6, 11]. Moreover, although we have previously shown reliable reporting of vaccination status in our study [6], our VE estimates and their relation with phylogenetic distance may be underestimated by nondifferential misclassification with regard to self-reported vaccination status. It remains to be seen whether cross-reactivity of 2vHPV following vaccination with <3 doses induces similar, long-lasting cross-protection. Analysis of the impact of 2vHPV among female teenagers in the United Kingdom shows evidence of type-specific protection, but not cross-protection following a single dose of vaccine [12].

One possibility as to why cross-protection is better explained by genomic distance than by measures based on the L1 capsid protein is that L1 VLP may induce cross-neutralizing antibodies to L2 that are critical in preventing viral entry into the host cell [1]. Moreover, the adjuvant AS04 in 2vHPV has been suggested to induce a T-cell response that enhances local innate control and provides help for subsequent adaptive immunity [13]. Although the mechanisms of adaptive immunity are still ambiguous, it is worthwhile to point out that vaccination with 2vHPV results not only in reduced incidence rates, but also in reduced viral load in breakthrough infections [14], suggesting that VE extends to control of infection postacquisition. AS04 is particularly effective in activating antigen-presenting cells, inducing cytokines and a T-helper 1–type response, leading to inhibition of viral transcription or translation [1]. Such features could contribute to cross-protective humoral and cellular control of HPV infections, and may be boosted by natural exposure to nonvaccine HPV types.

Cross-protection can be expected to mitigate the potential for type replacement by acting as a substitute of latent competitive pressures induced by vaccine types. Thus, oncogenic types that do not benefit from cross-reactivity should be considered foremost in evaluating type replacement in the wake of vaccination. In this context, the negative VE against HPV-59 might stem from differential sensitivity of the SPF10-LiPA25 assay in vaccinees relative to nonvaccinated controls [6], and not from type replacement. A Finnish community randomized trial on the population effects of 2vHPV, using a different PCR-based assay than ours, found no indications for type replacement by HPV-59 [15]. Instead, HPV-39 and HPV-51 were marked as potential culprits for an increased postvaccination occurrence.

To summarize, our analysis indicates that cross-protection from 2vHPV is sustained up to 8 years postvaccination and that the level of protection correlates with genomic distance to HPV-16 or -18. This suggests that the benefits of 2vHPV vaccination may extend to clinically relevant nonvaccine types, given that oncogenic potential of papillomaviruses itself has a phylogenetic basis [1–3]. Further studies will reveal to what extent cross-protection induced by the bivalent vaccine will contribute to HPV-related disease prevention.

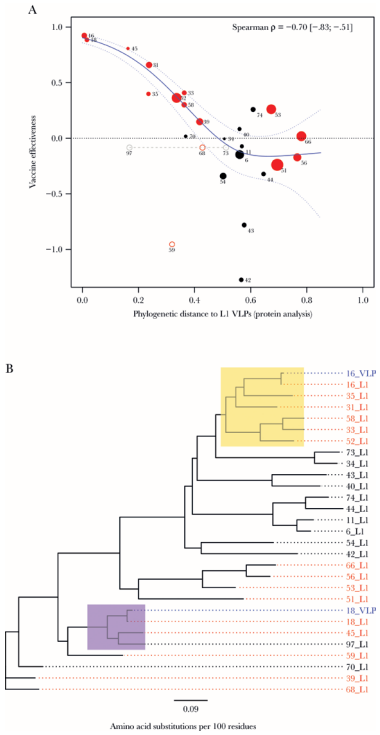
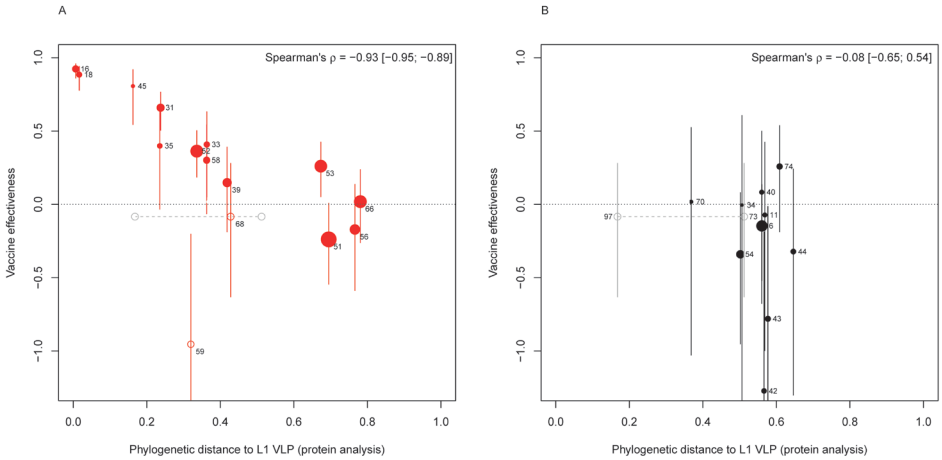


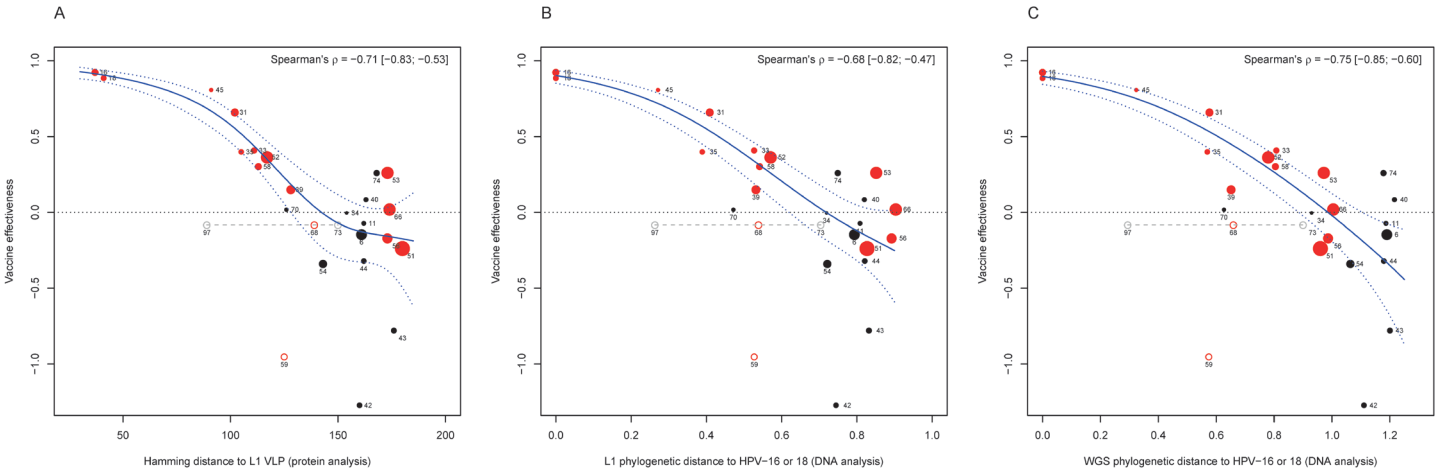
Figure 1. Bivalent human papillomavirus (HPV) vaccine effectiveness (VE) as a function of phylogenetic distance to L1 virus-like particles (VLPs). The VE was calculated from cross-sectional prevalence data [6] for all genotypes in the SPF10-LiPA25 assay. Phylogenetic distance to L1 VLPs was calculated from reference DNA sequences, using the Dayhoff model for evolutionary change in L1 protein. Substitution rate heterogeneity among alignment sites was incorporated by assuming γ-distributed rates plus a fraction of invariable sites. A, Red data points denote (possibly) oncogenic types, with HPV-68 being indistinguishable from HPV-73 and -97 (in gray). The size of each data point is plotted proportional to the number of positive test results (n) used in VE estimation. Spearman rank correlation (with 95% confidence limits in brackets) was calculated from all data points weighted by √n, excluding HPV-59 and -68 (open circles). The estimated spline function (in blue) is shown with 95% credible intervals (dotted lines). B, Phylogenetic tree based on L1 protein, with blue tips denoting reference sequences used for construction of L1 VLPs in the bivalent HPV vaccine and red tips denoting (possibly) oncogenic types in the SPF10-LiPA25 assay. The yellow and purple clades highlight types that are close enough to L1 VLPs to benefit (in principle) from cross-protection, according to predictions with 95% confidence from the weighted penalized regression spline in (A).

References

1. Roden RBS, Stern PL. Opportunities and challenges for human papillomavirus vaccination in cancer. Nat Rev Cancer 2018; 18:240–54.
2. de Villiers EM. Cross-roads in the classification of papillomaviruses. Virology 2013; 445:2–10.
3. Harper DM, DeMars LR. HPV vaccines—a review of the first decade. Gynecol Oncol 2017; 146:196–204.
4. Skinner SR, Apter D, De Carvalho N, et al. Human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine for the prevention of cervical cancer and HPV-related diseases. Expert Rev Vaccines 2016; 15:367–87.
5. Kavanagh K, Pollock KG, Cuschieri K, et al. Changes in the prevalence of human papillomavirus following a national bivalent human papillomavirus vaccination programme in Scotland: a 7-year cross-sectional study. Lancet Infect Dis 2017; 17:1293–302.
6. Woestenberg PJ, King AJ, van Benthem BHB, et al. Medical Microbiological Laboratories and the Public Health Services. Bivalent vaccine effectiveness against type-specific HPV positivity: evidence for cross-protection against oncogenic types among Dutch STI clinic visitors. J Infect Dis 2018; 217:213–22.
7. Mesher D, Panwar K, Thomas SL, Beddows S, Soldan K. Continuing reductions in HPV 16/18 in a population with high coverage of bivalent HPV vaccination in England: an ongoing cross-sectional study. BMJ Open 2016; 6:e009915.
8. Woestenberg PJ, King AJ, van der Sande MA, et al. Medical Microbiological Laboratories; Public Health Services. No evidence for cross-protection of the HPV-16/18 vaccine against HPV-6/11 positivity in female STI clinic visitors. J Infect 2017; 74:393–400.
9. Van Doorslaer K, Li Z, Xirasagar S, et al. The papillomavirus episteme: a major update to the papillomavirus sequence database. Nucleic Acids Res 2017; 45:D499–506.
10. Minh BQ, Nguyen MA, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol 2013; 30:1188–95.
11. Donken R, King AJ, Bogaards JA, Woestenberg PJ, Meijer CJLM, de Melker HE. High effectiveness of the bivalent human papillomavirus (HPV) vaccine against incident and persistent HPV infections up to 6 years after vaccination in young Dutch women. J Infect Dis 2018; 217:1579–89.
12. Cuschieri K, Kavanagh K, Moore C, Bhatia R, Love J, Pollock KG. Impact of partial bivalent HPV vaccination on vaccine-type infection: a population-based analysis. Br J Cancer 2016; 114:1261–4.
13. Didierlaurent AM, Morel S, Lockman L, et al. AS04, an aluminum salt- and TLR4 agonist-based adjuvant system, induces a transient localized innate immune response leading to enhanced adaptive immunity. J Immunol 2009; 183:6186–97.
14. van der Weele P, Breeuwsma M, Donken R, et al. Effect of the bivalent HPV vaccine on viral load of vaccine and non-vaccine HPV types in incident clearing and persistent infections in young Dutch females. PLoS One 2019; 14:e0212927.
15. Gray P, Palmroth J, Luostarinen T, et al. Evaluation of HPV type-replacement in unvaccinated and vaccinated adolescent females—post-hoc analysis of a community-randomized clinical trial (II). Int J Cancer 2018; 142:2491–500.



Supplementary Figure 1. Bivalent HPV vaccine effectiveness by oncogenicity versus phylogenetic distance to L1 VLP. Data points in A) and B) are similar to those in Fig. 1A but include 95% confidence intervals for VE. Spearman's rank correlation (with 95% confidence limits in brackets) in A) applies to (possibly) oncogenic types, excluding HPV-59 and 68, whereas the correlation in B) applies to non-oncogenic types, excluding HPV-73 and 97. The correlation in A) was ρ = -0.88 (95% CI: -0.93; -0.78) when also excluding HPV-16 and 18 (i.e. for non-vaccine types only).



Supplementary Figure 2. Bivalent HPV vaccine effectiveness as a function of alternative distance measures from HPV-16 and 18. In all panels, the size of each data point is plotted proportional to the number of positive test results n used in VE estimation, and Spearman's rank correlations (with 95% confidence limits in brackets) were calculated from all data points weighted by √n, excluding HPV-59 and 68 (open circles). Red data points denote (possibly) oncogenic types in the SPF10-LiPA25 assay (Table 1), with HPV-68 being indistinguishable from HPV-73 and 97 (in grey). Hamming distance in A) expresses the number of positions at which the L1 proteins of reference types have different amino acids than the VLPs in 2vHPV. Phylogenetic distance on L1 capsid genes in B) was calculated from reference DNA sequences, using the transversion model (AG=CT) and unequal base frequencies counted from the data [9]. Substitution rate heterogeneity among alignment sites was incorporated by assuming Gamma-distributed rates plus a fraction of invariable sites. Phylogenetic distance on whole genome sequences (WGS) in C) was calculated from reference DNA sequences, using the transversion model with AG=CT and empirical base frequencies [9]. Substitution rate heterogeneity among alignment sites was incorporated via a free rate model with five categories. The estimated spline functions (in blue) are shown with 95% credible intervals (dotted lines).

Supplementary Table 1. Bivalent HPV vaccine effectiveness (VE) against pooled estimates

N (%)		VE (95% CI) *		
		HPV-16/18	(possibly) oncogenic types ±	non-oncogenic types §
All women				
Unvaccinated	799 (38·0)	91·0 (86·0; 94·2)	25·8 (17·7; 33·2)	−4·9 (−20·7; 8·8)
Vaccinated (≥1 dose)	1305 (62·0)			
Women offered vaccination <5 years ago †				
Unvaccinated	185 (42·5)	85·4 (60·9; 94·5)	20·6 (−2·3; 38·4)	5·9 (−31·3; 32·6)
Vaccinated (≥1 dose)	250 (57·5)			
Women offered vaccination 5/6 years ago				
Unvaccinated	306 (37·7)	92·1 (83·9; 96·2)	30·6 (18·3; 41·3)	−14·2 (−42·8; 8·6)
Vaccinated (≥1 dose)	505 (62·3)			
Women offered vaccination 7/8 years ago				
Unvaccinated	308 (35·9)	91·7 (83·4; 95·8)	24·5 (11·4; 35·7)	−0·4 (−24·4; 18·9)
Vaccinated (≥1 dose)	550 (64·1)			

Abbreviations: HPV, human papillomavirus; 95% CI, 95% confidence interval
* VE was corrected for: age, migration background, education level, no. sex partners last 6 months, lifetime no. sex partners, age at sexual debut, history of sexually transmitted infections (STI), hormonal contraceptives use, STI-related symptoms and age vaccination was offered
± Including HPV types HPV-16/18/31/33/35/39/45/51/52/56/58/59/66/68 (73/97)
§ Including HPV types HPV-6/11/34/40/42/43/44/53/54/70/74
† Time since vaccination offered was calculated from the date of sampling relative to the scheduled date of vaccination, according to year of birth and the official schedule of HPV immunization in the Netherlands. For the catch-up cohorts, vaccination was offered on 01-03-2009. For the cohorts vaccinated routinely in the National Immunization Program, vaccination was offered on March 1 in the year they turn 13 year.

Supplementary Table 2. Rank correlation analysis between bivalent HPV vaccine effectiveness (VE) and various distance measures

Correlation coefficient	Measure				
		Hamming ±	**AA §**	**L1 †**	**WGS ††**
Spearman's ρ *	VE	-0·713 (p <0·001)	-0·703 (p <0·001)	-0·680 (p <0·001)	-0·753 (p <0·001)
	Hamming ±		0·963	0·957	0·744
	AA			0·959	0·712
	L1				0·765
Spearman's ρpartial ¶	VE	-0·158 (p =0·50)	-0·290 (p =0·20)	0·320 (p =0·15)	-0·512 (p =0·012)
	Hamming		0·429	0·414	0·046
	AA			0·571	-0·324
	L1				0·478
Spearman's $\rho_{partial}$	VE	(omitted)	-0·401 (p =0·063)	0·284 (p =0·21)	-0·526 (p =0·0087)
	AA			0·910	-0·337
	L1				0·546

Abbreviations: HPV, human papillomavirus; 95% CI, 95% confidence interval; AA, amino acid, WGS, whole genome sequence

* All correlation coefficients are weighted for the number of positive test results used in VE estimation (see Table 1)

¶ Partial correlation coefficients are obtained from the inverse weighted covariance matrix

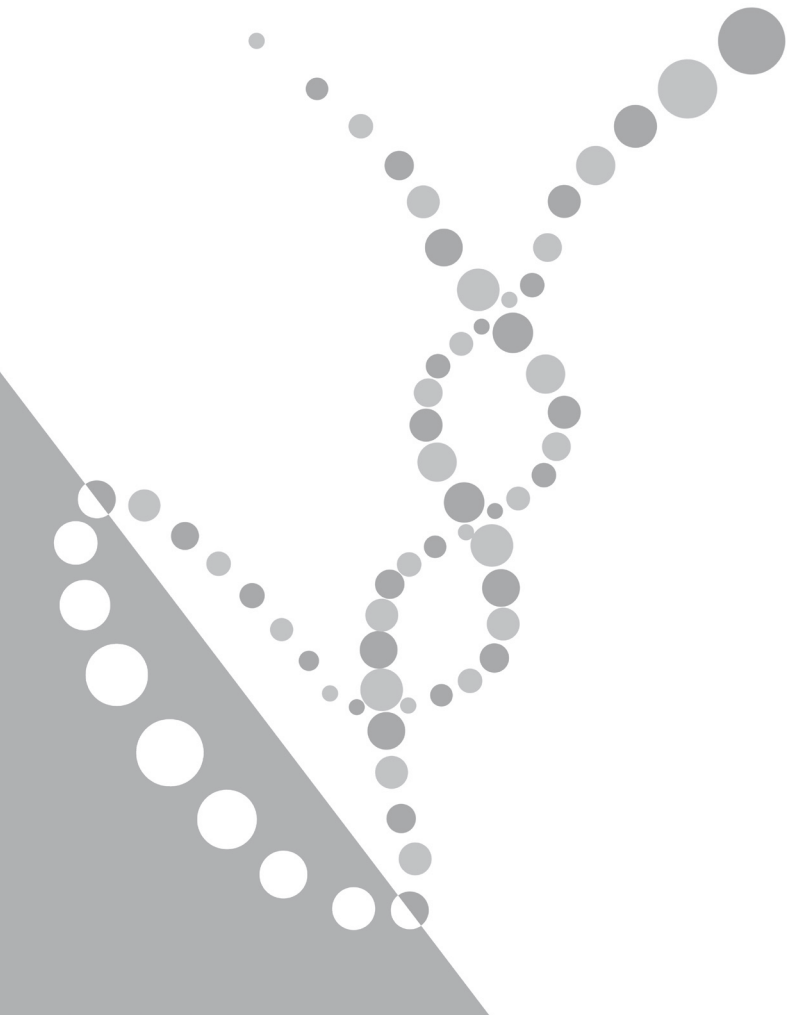
± Minimum Hamming distance between aligned L1 amino acid sequences of reference types and virus-like particles in bivalent vaccine

§ Minimum phylogenetic distance between aligned L1 amino acid sequences of reference types and bivalent vaccine types, HPV-16 and 18

† Minimum phylogenetic distance between aligned DNA sequences on the L1 capsid gene of reference types and bivalent vaccine types, HPV-16 and 18

†† Minimum phylogenetic distance between aligned DNA sequences on the whole viral genome of reference types and bivalent vaccine types, HPV-16 and 18

Chapter 9



General discussion

In 2009, the bivalent HPV vaccine was introduced into the National Immunization Program in the Netherlands. Parallel to its introduction, studies were initiated to monitor vaccine effectiveness and to gain in-depth insights into virological HPV infection characteristics pre- and post-vaccination. This thesis describes the potential value and relevance of molecular biomarkers in monitoring of HPV infections, specifically viral load measurement and whole genome sequencing. In the next part, we will highlight aspects that were not discussed in the previous chapters

Part I: Assessment of the value of viral load measurements

In the first part of this thesis, we set out to identify the value of viral load measurements in assessing persistent infections, in both non-vaccinated and vaccinated women. Trends in viral load measurements have been described numerously before. Until now, research focused mostly on adult women, describing trends and attempting to predict incident infection clearance, persistent infections and the development of different grades of CIN [1–5]. Viral load levels remained too volatile for use as a strong predictive biomarker, although limited clinical usefulness has been shown [6, 7]. There might be merit in repeated viral load measurements over time. Studies with extended follow-up have suggested a linear increase in viral load over time could be indicative of long-term disease progression [8, 9]. Currently, however, most studies use singular measurements, have limiting study sizes, or have limited follow-up, including the studies conducted in chapters 2 and 3. For clinical practice, ease of use is required and frequent follow-up over a long period is probably not feasible.

The limited availability of viral load data for young women as a potential biomarker for vaccine monitoring, inspired us to conduct viral load studies on age-appropriate, vaccine-eligible cohorts. Findings from these studies could then function as baseline data for vaccine monitoring studies. In chapter 2, we show that viral load trends described previously for mature women can also be applied to young, vaccine-eligible women. Infections that did not clear within a year were found to have a significantly higher median HPV16/18 viral load at baseline than infections that persisted for at least one year. However, the measurements were highly variable and as a result, linear regression analysis suggested poor predictive value for individuals. Consequentially, the value of viral load for detecting persistent infections could only be demonstrated in population or large cohort studies, such as vaccine monitoring studies.

We then moved on to the analysis of vaccine recipients in chapter 3. Viral loads of all hrHPV types, as well as HPV6 and HPV11 were measured in a cohort of both vaccinated and non-vaccinated young women. Significantly lower median viral load values were found in breakthrough infections of the vaccinated group for both HPV16 and HPV18 infections. Viral load reductions were also observed for non-vaccine types HPV6, 59 and 66, although this did not translate in lower incidence rates of infections. Furthermore, cross-protection was shown against some HPV types phylogenetically related to HPV16 and HPV18, but viral load measurements were volatile and did not align with reductions in type-specific incidence rates. To place our findings in context, we proposed a theory by which persistence in breakthrough (vaccine type) HPV infections could be inhibited.

The results in chapter 3 imply usefulness for viral load measurements in public health and epidemiological settings to monitor population-level (vaccine) effects. We show that the bivalent HPV vaccine is highly effective in reducing incidence of both clearing and persistent infections of vaccine type and phylogenetically related HPV types. The combined findings from chapters 2 and 3 suggest that viral load measurements have potential merit as a monitoring tool for measuring vaccine effectiveness, but that viral load on an individual basis, is not strong to predict a progressive infection. Therefore, viral load measurements are best used as a tool to support other epidemiological or molecular findings.

Currently, limited data is available about investigations of any such correlations. Identifying them would require large-scale studies with concerted measurements of multiple parameters. Fortunately, the method of viral load measurements, namely quantitative PCR, is relatively low-cost, potentially simple to multiplex and easy to perform in (automated) high-throughput. There could be merit in viral load measurements for public health, but it would require considerable effort to elucidate it.

Take home messages on HPV viral load measurements:

- Trends in HPV16 and HPV18 viral load measurements are similar for adult and young women
- Type specific HPV viral load measurements are too heterogeneous to reliably predict infection persistence at the individual level
- Type specific HPV viral load measurements can be used to assess population-level effects of prophylactic HPV vaccines, for both vaccine and non-vaccine HPV types
- Vaccination with the bivalent vaccine results in significantly lower HPV16 and HPV18 viral load breakthrough infections in vaccinated girls
- Viral load measurements alone are currently insufficient to unambiguously prove vaccine cross-protection against non-vaccine HPV types

Part II: Whole genome sequencing of HPV infections and its potential applications

The HPV genome sequence is associated with geographical heritage [10–12]. In addition, differences between HPV sequences, have been shown to affect infection persistence [13, 14], pathogenesis [15–19] and outcome of the infection [20, 21]. In the post-vaccination era, knowledge of HPV sequences against which the available vaccines convey protection is required. Sequencing of HPV genomes has expanded significantly over the past decade with the development of novel sequencing technologies, allowing for enhanced sensitivity of detection and exponentially increased read depth, facilitating detection of minor variants. In the second part of this thesis, we aimed to develop and apply pragmatic and novel sequencing techniques to elucidate HPV genome properties.

Prior to this thesis, whole genome HPV16 and HPV18 sequencing data was limited, especially for HPV18. In addition, the available data was focused largely on America and Asia, leaving sequences relevant to Europe to be desired [22–27]. In chapters 4 and 5, we investigate the whole

genome diversity level of HPV16 and HPV18 respectively in the same retrospective longitudinal cohort study used in chapter 2. For HPV16 we show in chapter 4 that sublineage A2 is dominant in the Netherlands, while only a limited number of A2 sequences had been described to date. In chapter 5, we show that for HPV18 sublineages A1 and A3 are dominant in the Netherlands, while prior studies from America and Asia predominantly identified sublineages A1 and B1. These findings highlight geographical trends in variant occurrence and imply validity of more global initiatives in sequencing of various (high-risk) HPV types. In addition, we have shown in chapters 4 and 5, that for both HPV16 and HPV18 strikingly diverse populations occur. This diversity is found despite strong conservation of the virus, which has been suggested to evolve at rates no more than ten times faster than the human genome [28, 29]. Almost every study participant was infected with a unique HPV16 or HPV18 variant and despite the observed diversity; sequences were entirely conserved in follow-up samples. Our findings suggest that transmission studies could be conducted for HPV lineage infections. This might be of interest in the context of public health, although a few caveats exist. For instance, variants need to remain conserved during transmission for the sake of traceability. The effect of transmission events on sequence stability is currently unknown. Logically the dominant variant in an infection is more likely to be transmitted, although no research has been conducted to support this. The transmission of HPV infections and their potential value for epidemiological transmission studies are potential areas of investigation for HPV research, which are now more easily facilitated by the availability of novel sequencing techniques. It should be noted, that despite our focus being predominantly on HPV16 and HPV18, sequencing of other HPV types might be of interest as well. While HPV16 and HPV18 are the most frequently occurring carcinogenic HPV types, they are by no means the most commonly circulating HPV types in the Netherlands or worldwide [30–33]. In the context of transmission studies, other more commonly circulating HPV types could be considered as a proof of concept, since very large studies would be required to evaluate transmission patterns of less common HPV types.

Additional public health relevance of chapters 4 and 5 is found in the rare events (< 5% of infections) where the major variant identified at baseline differed from the major variant at follow-up. NGS analysis of these infections showed that the major variant at follow-up was already present as a minority at baseline. In addition, the major variant at baseline was no longer detectable at follow-up, suggesting clearance of this specific variant. Considering the rarity of variant switches, conventional genotyping remains an adequate tool for identifying (persistent) infections in most cases. Variant detection could be relevant in the context of vaccine monitoring studies, where a single persistent infection in the vaccinated group could affect vaccine efficacy measurements in a negative fashion. In these cases, variant investigation could indicate if any breakthrough infection is truly persistent, or whether it consists of subsequent different incident infections.

The finding that HPV16 and HPV18 are strongly conserved over time, yet also highly diverse in a population is paradoxical and currently not well understood. Chapter 6 attempts to elucidate this finding by analyzing a subset of the samples used in chapter 4 with a highly sensitive next generation whole genome sequencing method called TaME-seq. HPV mutagenesis in high-grade lesions and cancer is suggested to be predominantly driven by the host innate anti-viral defense mechanism catalytic polypeptide-like apolipoprotein B mRNA editing enzyme (APOBEC), causing

C > T nucleotide shifts [18, 34–36]. In chapter 6, we suggest that there is another mechanism at play during earlier infection stages, since the mutation signatures from infections without clinical manifestation (predominantly T > C) are distinctly different from mutation signatures observed in high-grade CIN lesions and cervical cancer (predominantly C > T). This implies a mechanism at play, other than APOBEC, inducing T > C changes. Currently, the mutagenic process causing these changes is not well understood [34, 37] and future research will have to point out what causes these changes, what they mean, and how APOBEC changes become dominant over time.

Specific, potentially APOBEC related, variations in the HPV genome are associated with malignant outcomes [18], which could justify longitudinal tracking of these variants as they could carry an increased risk for transforming infections. In chapter 6, we show the presence of a large number of minority variants, which partially overlap with the CIN3 + associated minority variants described by Mirabello *et al*. The value of these minority variants is currently unknown, since it is also unknown how major variant sequence changes occur over time. However, using our method we could, in some cases, detect minority variants repeatedly in subsequent samples from the same infection. As the meaning of minority variants becomes clear, the value of variant tracking over time should become apparent for public health, in the form of screening programmes, and potentially for clinical practice, in the form of personalized medicine. However, although different HPV subtypes convey different cancer risks, these findings remain calculations of odds. Therefore, if any translation is made towards routine testing, a careful cost-benefit analysis is required to assess the potential benefits of sequencing analysis.

Conventional thinking suggests that the major variant is responsible for pathogenesis [38, 39], but we show in chapters 4 and 5 that major variants can switch over time. The data obtained in chapter 6, suggests a complex cocktail of single-nucleotide polymorphism (SNPs) being present at early stages of an HPV infection. The mutation signatures displayed by these infections grant insight in mutagenic processes to which a HPV infection is subjected, and by which minority variants are possibly generated. These SNPs indicate the presence of minority variants within a given HPV infection, although backtracking individual SNPs to biologically relevant variants is challenging. Novel sequencing platforms, such as nanopore sequencing, generate extremely long reads, potentially covering the HPV genome multiple times [40]. This could allow for investigation minority variants by inferring SNPs as parts of whole entities.

A clinical application of whole genome HPV sequencing

In chapter 7, we focus on women who were treated for CIN2 or CIN3, and who were diagnosed with recurrent CIN (rCIN) lesions after one to two years [41, 42]. When HPV16 was found in both the baseline CIN3 and rCIN, HPV16 variant analysis was performed. HPV16 variants identified at baseline CIN were almost completely conserved at follow-up, suggesting that the rCIN lesions were caused by the same HPV16 variant that caused the initial CIN2/3 lesion. The fact that the same variant is still present post-treatment and able to cause new lesions within 6–12 months suggests incomplete or inadequate ablative treatment of the CIN2/3 lesion or that women are unable to clear these specific variants. Our results show that in virtually all cases where the same

HPV type is found in both baseline CIN and rCIN, the same HPV variant is found. This implies that routine sequencing of these lesions has no additional value for clinical practice and that conventional HPV genotyping suffices.

Potential sequencing research for the future

The obvious next step in HPV sequencing research, is analyzing breakthrough infections in vaccinated individuals. Initially for HPV16 and HPV18, and eventually for all high-risk HPV types. Touching back on the cross-protection described in chapter 3, it would be of interest to investigate HPV types against which some form of protection has been described at the sequence level. Cross-protection of the bivalent vaccine is probably caused by a relatively broad immune response against phylogenetically related HPV types facilitated by virus-like particle composition of the vaccine and the AS04 adjuvant [43, 44]. However, it has also been previously suggested that the partial cross-protection against HPV31 is limited to specific lineages [45], with other lineages being relatively unaffected by vaccination. The combination of these findings suggests a mechanism in which the vaccine not only has type-specific vaccine effectiveness, but perhaps also lineage (or even variant) specific effectiveness. The methods developed and studies used in this thesis, allow for an investigation of HPV infections in vaccinated individuals and a subsequent comparison to available data. This would allow us to investigate potential genetic changes in lineages or variants of specific HPV types driven by vaccination. Previous research has suggested that genetic change of HPV following HPV vaccination is improbable, but possible, although it has never been demonstrated directly [46–49]. The sequencing of non-vaccine types would grant insight in the cross-protective effectiveness of the bivalent HPV vaccine against non-vaccine HPV types and its possible limitations in protecting against specific lineages of certain HPV types.

Non-targeted alternatives for HPV detection

In this thesis, we have made use of a number of targeted HPV sequencing approaches. In chapters 4, 5 and 7, we use various PCR approaches followed by Sanger or Illumina sequencing. In chapter 6, we apply TaME-seq, an adapted Nextera library preparation method in combination with Illumina sequencing. These targeted approaches are often pragmatic, making detection of low copy numbers feasible. However, targeting specific sequences using a selection of primers, introduces bias towards variants that match those primers as described in the introduction of this thesis. Consequentially, in our studies we might underestimate viral diversity, as we are unable to detect variants against which our primers do not fit. The alternative would be unbiased, random amplification of DNA, possibly in combination with depletion of unwanted sources of DNA or semi-targeted approaches enriching viral DNA specifically using probe enrichment methods [50–52], sequence-independent, single-primer amplification (SISPA) [53–55], or rolling-circle amplification [56–58]. These methods can be used on samples while making no assumptions for their contents and for virus discovery. The downside of semi-targeted and non-targeted sequencing approaches is the generally lower ratio of HPV sequence reads compared to off-target reads, requiring deeper sequencing.

The cross-roads of epidemiology and sequencing

In chapter 8 we attempted to model vaccine efficiency against various HPV types by combining reference sequence data with actual vaccine efficiency data obtained from a Dutch study. The results coincide largely with those reported by previous vaccine efficacy trials and monitoring studies, but expand on them by including other non-vaccine HPV types. The results of our modeling study indicate that vaccine effectiveness is strongly correlated with phylogenetic distance between HPV types and the virus-like particles included in the bivalent vaccine. The outcome of this study functions as a proof of concept and paves the way for larger scale studies, and perhaps even meta-analyses.

Parallel to the in-depth investigation of the bivalent vaccine, cross-protection could be modeled for the quadrivalent and potentially the nonavalent vaccine, by applying the same modeling technique to vaccine-specific cohort studies and vaccine-specific virus-like particle DNA sequences.

By combining results from multiple monitoring studies, more reliable protection estimates could be modeled. Ideally, protection estimates could even be modeled in a more relevant fashion by including local or global sequencing results. In theory, this is already possible using repository sequence data, along with vaccine monitoring study data if they are made available as well. In doing so, (sub) lineage specific protection estimates might be calculated for HPV types against which the protection is low and thus measurable only in very large studies. This could be relevant in assessing future screening targets, when the usual suspects HPV16 and HPV18 are less of a threat thanks to vaccination.

Take home messages on HPV whole genome sequencing:

- HPV is well-suited for targeted sequencing approaches due to its conserved double-stranded DNA composition
- Agnostic sequencing approaches might reduce bias for known HPV types and variants
- HPV16 and HPV18 comprise a highly diverse population of major variants that are strongly conserved through time
- HPV type sequence diversity is geographically dependent
- High resolution sequencing of HPV16 has shown a plethora of minority variants present in infections, of which the exact meaning is currently unknown
- The mutation signatures of pre-clinical infections seem distinctly different from those of infection with cervical lesions
- DNA sequencing data of HPV16 positive CIN2/3 at baseline and HPV16 positive recurrent CIN after treatment, has shown that in the large majority of cases identical variants cause rCIN
- HPV genome sequence correlates well with vaccine effectiveness estimates and might help reconcile different cross-protective effects described by various studies

Conclusion

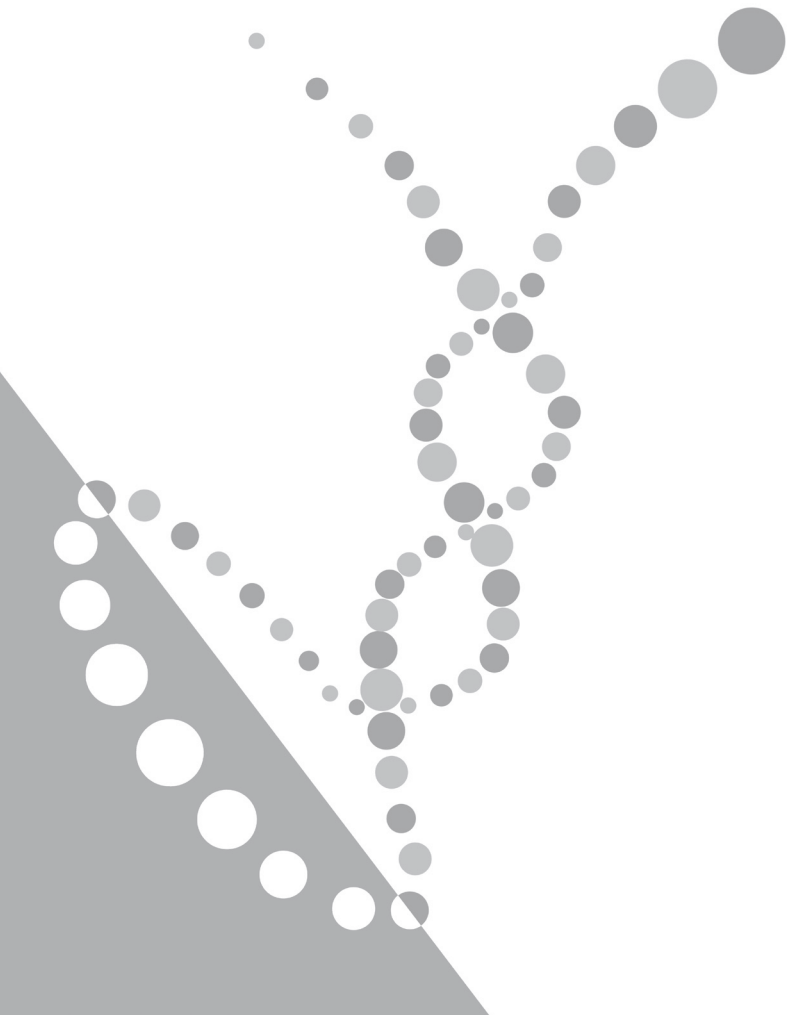
HPV infection is a complex interplay of elements at the molecular level and markers for persistent infections can give early insight in infection development. Using viral load measurements, we were able to reinforce the high efficiency of bivalent HPV vaccination on the population level. Through various methods of sequencing, HPV was shown to be strongly conserved, yet highly diverse at the same time. These findings could be used for epidemiological transmission studies. In addition, the monitoring of minority variants over time is possible, although clinical relevance of this type of monitoring has yet to be shown. Furthermore, the longitudinal analysis of majority variants has shown that if a CIN lesion and a post-treatment recurrent CIN lesion are both caused by HPV16, they are most likely caused by the same initial HPV16 infection. The combination of epidemiological and molecular data allowed for modeling of vaccine efficiency, which could aid the public health effort by implicating new potential targets for vaccination, monitoring or screening initiatives.

References

1. Briolat, J., et al., *HPV prevalence, viral load and physical state of HPV-16 in cervical smears of patients with different grades of CIN.* Int J Cancer, 2007. **121**(10): p. 2198–204.
2. Winer, R.L., et al., *Quantitative human papillomavirus 16 and 18 levels in incident infections and cervical lesion development.* J Med Virol, 2009. **81**(4): p. 713–21.
3. Winer, R.L., et al., *Viral load and short-term natural history of type-specific oncogenic human papillomavirus infections in a high-risk cohort of midadult women.* Int J Cancer, 2014. **134**(8): p. 1889–98.
4. Franco, E.L. and F. Coutlee, *Prognostic value of measuring load of human papillomavirus DNA in cervical samples: an elusive target.* J Natl Cancer Inst, 2009. **101**(3): p. 131–3.
5. Trevisan, A., et al., *Human papillomavirus type 16 viral load measurement as a predictor of infection clearance.* J Gen Virol, 2013. **94**(Pt 8): p. 1850–7.
6. Hesselink, A.T., et al., *High-risk human papillomavirus DNA load in a population-based cervical screening cohort in relation to the detection of high-grade cervical intraepithelial neoplasia and cervical cancer.* Int J Cancer, 2009. **124**(2): p. 381–6.
7. Snijders, P.J., et al., *Determination of viral load thresholds in cervical scrapings to rule out CIN 3 in HPV 16, 18, 31 and 33-positive women with normal cytology.* Int J Cancer, 2006. **119**(5): p. 1102–7.
8. Depuydt, C.E., et al., *Changes in type-specific human papillomavirus load predict progression to cervical cancer.* J Cell Mol Med, 2012. **16**(12): p. 3096–104.
9. Depuydt, C.E., et al., *Serial type-specific human papillomavirus (HPV) load measurement allows differentiation between regressing cervical lesions and serial virion productive transient infections.* Cancer Med, 2015. **4**(8): p. 1294–302.
10. Ho, L., et al., *The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations.* J Virol, 1993. **67**(11): p. 6413–23.
11. Yamada, T., et al., *Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective.* J Virol, 1997. **71**(3): p. 2463–72.
12. Pimenoff, V.N., C.M. de Oliveira, and I.G. Bravo, *Transmission between Archaic and Modern Human Ancestors during the Evolution of the Oncogenic Human Papillomavirus 16.* Mol Biol Evol, 2017. **34**(1): p. 4–19.
13. Xi, L.F., et al., *Human papillomavirus type 16 and 18 variants: race-related distribution and persistence.* J Natl Cancer Inst, 2006. **98**(15): p. 1045–52.
14. Zhang, L., et al., *Variants of human papillomavirus type 16 predispose toward persistent infection.* Int J Clin Exp Pathol, 2015. **8**(7): p. 8453–9.
15. Andersson, S., et al., *Uneven distribution of HPV 16 E6 prototype and variant (L83V) oncoprotein in cervical neoplastic lesions.* Br J Cancer, 2000. **83**(3): p. 307–10.
16. Sichero, L., J.S. Sobrinho, and L.L. Villa, *Oncogenic potential diverge among human papillomavirus type 16 natural variants.* Virology, 2012. **432**(1): p. 127–32.
17. Mirabello, L., et al., *HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women.* J Natl Cancer Inst, 2016. **108**(9).
18. Mirabello, L., et al., *HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis.* Cell, 2017. **170**(6): p. 1164–1174 e6.
19. Xi, L.F., et al., *Risk for high-grade cervical intraepithelial neoplasia associated with variants of human papillomavirus types 16 and 18.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(1): p. 4–10.

20. Lizano, M., et al., *Distribution of HPV16 and 18 intratypic variants in normal cytology, intraepithelial lesions, and cervical cancer in a Mexican population.* Gynecol Oncol, 2006. **102**(2): p. 230–5.
21. Cullen, M., et al., *Deep sequencing of HPV16 genomes: A new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection.* Papillomavirus Research, 2015.
22. Chen, Z., et al., *Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses.* J Virol, 2005. **79**(11): p. 7014–23.
23. Lurchachaiwong, W., et al., *Entire genome characterization of human papillomavirus type 16 from infected Thai women with different cytological findings.* Virus Genes, 2009. **39**(1): p. 30–8.
24. Sun, M., et al., *Whole genome sequencing and evolutionary analysis of human papillomavirus type 16 in central China.* PLoS One, 2012. **7**(5): p. e36577.
25. de Oliveira, C.M., et al., *High-level of viral genomic diversity in cervical cancers: A Brazilian study on human papillomavirus type 16.* Infect Genet Evol, 2015. **34**: p. 44–51.
26. Lurchachaiwong, W., et al., *Whole-genome sequence analysis of human papillomavirus type 18 from infected Thai women.* Intervirology, 2010. **53**(3): p. 161–6.
27. Chen, Z., et al., *Evolution and taxonomic classification of alphapapillomavirus 7 complete genomes: HPV18, HPV39, HPV45, HPV59, HPV68 and HPV70.* PLoS One, 2013. **8**(8): p. e72565.
28. Bravo, I.G. and M. Felez-Sanchez, *Papillomaviruses: Viral evolution, cancer and evolutionary medicine.* Evol Med Public Health, 2015. **2015**(1): p. 32–51.
29. Van Doorslaer, K., *Evolution of the papillomaviridae.* Virology, 2013. **445**(1–2): p. 11–20.
30. Donken, R., et al., *High effectiveness of the bivalent HPV vaccine up to six years post-vaccination against incident and persistent HPV infections in young Dutch females.* J Infect Dis, 2018. **217**(10): p. 1579–1589.
31. Woestenberg, P.J., et al., *Bivalent Vaccine Effectiveness Against Type-Specific HPV Positivity: Evidence for Cross-Protection Against Oncogenic Types Among Dutch STI Clinic Visitors.* J Infect Dis, 2018. **217**(2): p. 213–222.
32. de Sanjose, S., et al., *Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis.* Lancet Infect Dis, 2007. **7**(7): p. 453–9.
33. de Sanjose, S., et al., *Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study.* Lancet Oncol, 2010. **11**(11): p. 1048–56.
34. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415–21.
35. Hirose, Y., et al., *Within-Host Variations of Human Papillomavirus Reveal APOBEC Signature Mutagenesis in the Viral Genome.* J Virol, 2018. **92**(12).
36. Wallace, N.A. and K. Munger, *The curious case of APOBEC3 activation by cancer-associated human papillomaviruses.* PLoS Pathog, 2018. **14**(1): p. e1006717.
37. Harris, R.S., *Cancer mutation signatures, DNA damage mechanisms, and potential clinical implications.* Genome Med, 2013. **5**(9): p. 87.
38. Xi, L.F., et al., *Human papillomavirus type 16 variants in paired enrollment and follow-up cervical samples: implications for a proper understanding of type-specific persistent infections.* J Infect Dis, 2010. **202**(11): p. 1667–70.
39. Xi, L.F., et al., *Analysis of human papillomavirus type 16 variants indicates establishment of persistent infection.* J Infect Dis, 1995. **172**(3): p. 747–55.
40. Payne, A., et al., *BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files.* Bioinformatics, 2018.

41. Polman, N.J., et al., *Good performance of p16/ki-67 dual-stained cytology for surveillance of women treated for high-grade CIN.* Int J Cancer, 2017. **140**(2): p. 423–430.
42. Uijterwaal, M.H., et al., *Performance of CADM1/MAL-methylation analysis for monitoring of women treated for high-grade CIN.* Gynecol Oncol, 2016. **143**(1): p. 135–142.
43. Harper, D.M. and L.R. DeMars, *HPV vaccines–A review of the first decade.* Gynecol Oncol, 2017. **146**(1): p. 196–204.
44. Schiller, J. and D. Lowy, *Explanations for the high potency of HPV prophylactic vaccines.* Vaccine, 2018.
45. Harari, A., et al., *Cross-protection of the Bivalent Human Papillomavirus (HPV) Vaccine Against Variants of Genetically Related High-Risk HPV Infections.* J Infect Dis, 2016. **213**(6): p. 939–47.
46. Gray, P., et al., *Evaluation of HPV type-replacement in unvaccinated and vaccinated adolescent females-Post-hoc analysis of a community-randomized clinical trial (II).* Int J Cancer, 2018. **142**(12): p. 2491–2500.
47. Mesher, D., et al., *Population-Level Effects of Human Papillomavirus Vaccination Programs on Infections with Nonvaccine Genotypes.* Emerg Infect Dis, 2016. **22**(10): p. 1732–40.
48. Murall, C.L., K.S. McCann, and C.T. Bauch, *Revising ecological assumptions about Human papillomavirus interactions and type replacement.* J Theor Biol, 2014. **350**: p. 98–109.
49. Tota, J.E., et al., *Epidemiologic approaches to evaluating the potential for human papillomavirus type replacement postvaccination.* Am J Epidemiol, 2013. **178**(4): p. 625–34.
50. Sunshine, S., et al., *HIV Integration Site Analysis of Cellular Models of HIV Latency with a Probe-Enriched Next-Generation Sequencing Assay.* J Virol, 2016. **90**(9): p. 4511–4519.
51. O'Flaherty, B.M., et al., *Comprehensive viral enrichment enables sensitive respiratory virus genomic identification and analysis by next generation sequencing.* Genome Res, 2018. **28**(6): p. 869–877.
52. Xiao, Y., et al., *Design and validation of a universal influenza virus enrichment probe set and its utility in deep sequence analysis of primary cloacal swab surveillance samples of wild birds.* Virology, 2018. **524**: p. 182–191.
53. Reyes, G.R. and J.P. Kim, *Sequence-independent, single-primer amplification (SISPA) of complex DNA populations.* Mol Cell Probes, 1991. **5**(6): p. 473–81.
54. Cholleti, H., et al., *Genetic characterization of a novel picorna-like virus in Culex spp. mosquitoes from Mozambique.* Virol J, 2018. **15**(1): p. 71.
55. Ni, J., et al., *Identification and genomic characterization of a novel porcine parvovirus (PPV6) in China.* Virol J, 2014. **11**: p. 203.
56. Rector, A., R. Tachezy, and M. Van Ranst, *A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification.* J Virol, 2004. **78**(10): p. 4993–8.
57. Latsuzbaia, A., et al., *Complete Genome Sequence of a Novel Human Gammapapillomavirus Isolated from a Cervical Swab in Luxembourg.* Genome Announc, 2018. **6**(11).
58. Pastrana, D.V., et al., *Metagenomic Discovery of 83 New Human Papillomavirus Types in Patients with Immunodeficiency.* mSphere, 2018. **3**(6).



Appendix

Summary

Human papillomavirus (HPV) is the most common sexually transmitted infection worldwide. Most HPV infections are benign and are cleared by the host within six to twelve months, although some infections can persist. A persistent infection with an oncogenic HPV type is required for the development of cervical cancer. In addition, oncogenic HPV type infections can lead to a number of other cancers in both females and males. The development from initial infection to cervical cancer can take decades and progresses through histopathologically well-defined pre-cancerous stages called cervical intraepithelial neoplasia 1–3 (CIN1–3).

Over 200 different HPV types have been described, of which only a limited number are capable of causing cervical cancer. These HPV types are considered oncogenic HPV types. Currently, three highly efficacious prophylactic vaccines are available to prevent HPV-related cervical disease. The bivalent vaccine protects against HPV16 and HPV18, potentially preventing up to 70% of all cervical cancer cases. The quadrivalent vaccine also protects against HPV16 and HPV18, but further includes protection against HPV6 and HPV11, thereby preventing most cases of genital warts. The more recently introduced nonavalent vaccine further extends protection of the quadrivalent vaccine by including HPV31, 33, 45, 52 and 58, which cause approximately 20% of all cervical cancer cases. All three vaccines convey strong protection against the targeted HPV types. Cross-protection against other, non-vaccine HPV types has been described for all vaccines, but has been found strongest for the bivalent vaccine.

In the Netherlands, prophylactic vaccination against cervical cancer was initiated in 2009 using the bivalent vaccine. Parallel to its introduction, a vaccine monitoring programme was initiated, to monitor vaccine effects. The earliest conventional marker to monitor vaccine effectiveness is the presence of persistent infections in vaccinated individuals. This thesis explores the value of HPV viral load measurements to assess (persistent) infections in both vaccinated and non-vaccinated individuals. In addition, long-term HPV vaccination could lead to selective pressure on the HPV population, potentially leading to changes in the prevalence of specific HPV genome variants. In this thesis, we attempt to gain insight into such events, by developing HPV whole genome sequencing assays and applying them to relevant cohort studies.

Part I: Viral load measurements in vaccinated and non-vaccinated settings

The value of viral load measurements is assessed in **part I** of this thesis. In **chapter 2**, we analyzed HPV16 and HPV18 viral load measurements obtained from a study of young, vaccine-eligible women (aged 16–29 years old). The viral load trends we identified from these young women were similar to trends described in previous studies including mature women. Baseline HPV16 and HPV18 viral load measurements were found to be significantly higher at the population level for infections that persist for at least one year compared to infections that clear within a year. However, these differences were not strong enough to generate discriminatory value between clearing and persistent infections at the individual level. Consequently, these viral load assays were used in order to analyze possible population-level effects of the bivalent HPV vaccine on viral load in **chapter 3**. Here, viral load measurements were conducted for all oncogenic HPV types as well as HPV6 and HPV11. For the vaccine types HPV16 and HPV18, significantly lower viral loads were found in both breakthrough incident clearing and incident persistent infections in vaccinated individuals compared to infections in non-vaccinated individuals. This finding implies that breakthrough HPV16 and HPV18 infections in vaccinated individuals are potentially less likely to persist and cause cervical disease. Cross-protection was found against HPV31, 33, 35 and 45, but these effects could not be correlated with viral load measurements.

Part II: Development and application of (next-generation) sequencing assays in epidemiological and clinical contexts

In **part II** of this thesis, we explore HPV16 and HPV18 sequence diversity in pre-vaccination and clinical studies. In **chapters 4 and 5**, both persistent and clearing HPV16 and HPV18 infections were studied in a cohort of unvaccinated women. The resulting HPV16 and HPV18 populations were found to be remarkably diverse considering the conserved nature of the virus. Nearly all study participants were found to be infected with unique HPV16 or HPV18 variants, potentially opening research opportunities for transmission studies. In persistent infections, the HPV sequence was found to be strongly conserved over the three-year follow-up period of the study. In rare cases, the variant sequenced at baseline did not match with the variant sequenced at follow-up. This implies that over time, variant switching may occur. When persistent HPV type-specific infections are identified in vaccinated women, infection with a different variant of the same HPV type should be excluded. This finding may be relevant in vaccine monitoring studies.

The switching of variants over time raises the possibility that multiple HPV variants of the same type could be present at any one time. The presence of these so-called minority variants was explored in **chapter 6**. A subset of the same study used in **chapters 4 and 5** was used to study HPV16 intra-host diversity. Using a novel, highly sensitive next-generation sequencing assay, a large number of intra-host minority variants were identified. The mutation patterns of these infections, identified in young women, appeared to be different when compared to mutation patterns of CIN2/3 infections. These differences potentially suggest that mutation processes

identified in HPV present in CIN2/3 lesions are not (yet) present in HPV infections not associated with CIN. These mutational signatures could be an indication of the biological processes driving HPV diversity generation.

In **chapter 7**, we assessed HPV16 variants from HPV 16 positive women who were treated for CIN2/3 and who were found to have HPV16 positive recurrent CIN (rCIN) during follow-up. Usually, the majority (90%-95%) of women who undergo ablative treatment (LLETZ or LEEP) after CIN2/3 detection are able to clear the infection. However, a minority of women (5%-10%) contract rCIN. Sequencing of both the baseline HPV16-associated CIN and rCIN samples one to two years later resulted in (nearly) identical sequences before and after treatment. The presence of the same HPV16 variant in both baseline and posttreatment rCIN suggests that the treatment procedure was inadequate or that these women are unable to clear these specific variants.

Vaccine cross-protection is currently being studied extensively, and although there is consensus that HPV vaccines can cause cross-protection against other non-vaccine HPV types, discrepancies between studies do exist. In **chapter 8** of this thesis, we correlate HPV reference sequence data with HPV vaccine effectiveness data. Phylogenetic distance of a given HPV type to HPV16 and HPV18 vaccine protein sequence is shown to strongly correlate with vaccine effectiveness against that HPV type. Stronger cross-protection was inferred against HPV types phylogenetically closely related to vaccine HPV types. On the other hand, cross-protection against HPV types phylogenetically distant from the vaccine HPV types is unlikely. These findings might help to reconcile differences in findings between vaccine monitoring studies.

The studies described in this thesis have resulted in new methods to monitor the presence of persistent HPV infections in vaccinated women. Using viral load measurements, we were able to reinforce the high vaccine efficiency of bivalent HPV vaccination at the population level. Through various methods of sequencing, HPV was shown to be strongly conserved over time, yet highly diverse at the same time. Furthermore, HPV16/18 variant analysis may be useful to exclude repeated infections with different variants of the same HPV type in breakthrough infections in vaccinated women.

Nederlandse samenvatting

Humaan papillomavirus is de meest voorkomende seksueel overdraagbare infectie wereldwijd. De meeste HPV infecties zijn goedaardig en zullen door de gastheer geklaard worden na zes tot twaalf maanden, echter, sommige infecties kunnen persisteren. Een persistente (langdurige) infectie met een oncogeen HPV type is noodzakelijk voor het ontwikkelen van baarmoederhalskanker. Naast baarmoederhalskanker kunnen persistente infecties met een oncogeen HPV type ook leiden tot een aantal andere vormen van kanker, zowel bij vrouwen als bij mannen. De ontwikkeling van beginnende infectie tot baarmoederhalskanker duurt decennia en gedurende dit proces worden een aantal histopathologisch gedefinieerde voorstadia van kanker doorlopen, genaamd cervicaal intra-epitheliale neoplasie 1–3 (CIN1–3).

Tot nu toe zijn er meer dan 200 verschillende HPV types ontdekt en beschreven, waarvan slechts een zeer beperkt aantal in staat is baarmoederhalskanker te veroorzaken. Deze types worden oncogene HPV types genoemd. Momenteel zijn er drie zeer effectieve preventieve vaccins beschikbaar die in staat zijn HPV-gerelateerde baarmoederhalsaandoeningen te voorkomen. Ten eerste het bivalente vaccin. Dit vaccin beschermt tegen HPV16 en HPV18, en voorkomt zo tot 70% van alle baarmoederhalskankergevallen. Het quadrivalente vaccin beschermt ook tegen HPV16 en HPV18, maar voegt bescherming tegen HPV6 en HPV11 toe. Deze types veroorzaken nagenoeg alle gevallen van genitale wratten. Recentelijk is het nonavalente vaccin geïntroduceerd. Dit vaccin breidt de bescherming van het quadrivalente vaccin uit met bescherming tegen HPV31, 33, 35, 45 en 52, welke ongeveer 20% van alle baarmoederhalskankergevallen veroorzaken. Alle drie de vaccins beschermen zeer goed tegen de types waartegen ze gericht zijn. Daarnaast is voor alle drie de vaccins ook kruisbescherming aangetoond tegen andere HPV types waartegen de vaccins niet gericht zijn. Dit effect is het sterkst gevonden voor het bivalente vaccin.

In Nederland wordt sinds 2009 gevaccineerd met het bivalente HPV vaccin als onderdeel van het Rijksvaccinatieprogramma om baarmoederhalskanker te voorkomen.

Tegelijkertijd werd er, naast introductie van het vaccin, een programma gestart om vaccineffecten te monitoren. Aangezien de transitie van initiële infectie tot baarmoederhalskanker een langdurig proces is, moet gebruik worden gemaakt van intermediaire eindpunten om de werking van het vaccin te evalueren. De vroegst mogelijk bruikbare indicator is die van persisterende infecties in gevaccineerde individuen. In dit proefschrift onderzoeken we de waarde van HPV viral load bepalingen (kwantificatie van de hoeveelheid virus) om inzicht te krijgen in persisterende HPV infecties in zowel gevaccineerde als ongevaccineerde individuen. Tevens ontwikkelen we in dit proefschrift technieken om het genoom van HPV eenvoudig en volledig in beeld te krijgen,

aangezien op de lange termijn HPV vaccinatie mogelijk kan leiden tot selectieve druk op de HPV populatie. Dit zou dan kunnen leiden tot veranderingen in het voorkomen van specifieke HPV genoomvarianten. Zowel de viral load bepalingen als het in kaart brengen van het HPV genoom zijn na ontwikkeling toegepast op relevante cohortstudies om inzicht in vaccineffecten te verkrijgen.

Deel I: Viral load bepalingen in gevaccineerde en ongevaccineerde settings

De waarde van HPV viral load bepaling wordt uiteengezet in **deel I** van dit proefschrift. In **hoofdstuk 2** hebben we HPV16 en HPV18 gekwantificeerd in een studie onder jonge vrouwen (16–29 jaar oud) die in aanmerking zouden komen voor vaccinatie in het huidige vaccinatieschema. De trends die zijn gevonden qua hoeveelheid virus in infecties in deze jonge vrouwen, waren vergelijkbaar met trends beschreven in voorgaand onderzoek onder oudere vrouwen. De hoeveelheid HPV16 en HPV18 in infecties op het eerste meetmoment bleek op populatieniveau significant hoger te zijn bij infecties die ten minste één jaar persisteren dan bij infecties die binnen één jaar geklaard werden. Deze verschillen waren echter niet sterk genoeg om voorspellende waarde te generen voor individuele persisterende of klarende infecties.

De in **hoofdstuk 2** ontwikkelde viral load bepalingen zijn in **hoofdstuk 3** gebruikt om op populatieniveau vaccineffecten te analyseren. In **hoofdstuk 3** zijn naast metingen aan HPV16 en HPV18 nieuwe testen geïntroduceerd om ook de hoeveelheid virus te kwantificeren in infecties met alle andere oncogene HPV types, alsmede HPV6 en HPV11 infecties. Voor de vaccintypes HPV16 en HPV18 vonden we significant lagere hoeveelheden virus in incident klarende en persisterende doorbraakinfecties in gevaccineerde studiedeelneemsters dan in ongevaccineerde deelneemsters. Dit betekent dat mocht een gevaccineerd persoon toch een doorbraak HPV16 of HPV18 infectie krijgen, dat deze infecties waarschijnlijk niet zullen persisteren en geen cervicale aandoeningen kunnen veroorzaken. Naast deze bevinding werd ook kruisbescherming van het bivalente vaccin tegen HPV31, 33, 35 en 45 gevonden, al kon dit niet gecorreleerd worden met metingen aan de hoeveelheid virus in infecties.

Deel II: Ontwikkeling en toepassing van (next-generation) sequentiebepalingen in epidemiologische en klinische context

In **deel II** van dit proefschrift onderzoeken we genoomdiversiteit van HPV16 en HPV18 in pre-vaccinatie studies en klinische studies. In **hoofdstukken 4 en 5** werd specifiek gekeken naar de sequentie van HPV16 en HPV18 infecties in een cohortstudie onder ongevaccineerde vrouwen. De resulterende HPV16 en HPV18 populaties waren zeer divers, ondanks de sterke conservering van het viraal genomisch DNA. Nagenoeg alle studiedeelneemsters waren geïnfecteerd met een unieke HPV16 of HPV18 variant, maar in persisterende infecties bleef de sequentie volledig behouden gedurende de drie jaar dat vrouwen aan deze studie deelnamen. In enkele zeldzame gevallen observeerden we op opvolgmomenten een andere virusvariant dan aan het begin van de studie. Dit impliceert dat er over tijd herinfecties met andere varianten plaats kunnen vinden.

Wanneer een persisterende infectie wordt waargenomen in een gevaccineerd persoon, moet uitgesloten worden dat het een herinfectie met een andere virusvariant betreft. Dit kan namelijk negatieve effecten hebben voor vaccinmonitoring. In HPV vaccinstudies wordt vaccineffectiviteit onder andere gemeten tegen het aantal persistente infecties dat wordt waargenomen binnen de studie. Wanneer een persistente infectie wordt waargenomen, terwijl het eigenlijk een herinfectie met een andere variant van hetzelfde HPV type betreft, leidt dit ten onrechte tot rapportage van een lagere vaccineffectiviteit.

Het wisselen van varianten betekent mogelijk ook dat meerdere varianten van hetzelfde HPV type op hetzelfde moment aanwezig kunnen zijn. De aanwezigheid van deze zogenoemde minderheidsvarianten werd onderzocht in **hoofdstuk 6**. Hier werd een deel van de monsters van de studiepopulatie van **hoofdstuk 4 en 5** gebruikt om HPV16 diversiteit binnen de gastheer te inventariseren. Aan de hand van een nieuwe, zeer gevoelige methode voor sequentiebepaling werd een groot aantal minderheidsvarianten geïdentificeerd. De mutatiepatronen die gevonden werden voor jonge vrouwen, bleken wezenlijk te verschillen van mutatiepatronen die geassocieerd zijn met hooggradige CIN laesies. Deze verschillen wijzen er mogelijk op dat mutatie-inducerende processen in hooggradige CIN laesies anders zijn dan de processen in infecties die niet geassocieerd zijn met CIN laesies. Het door ons gevonden mutatiepatroon kan ook een indicator zijn van het biologisch proces dat ten grondslag ligt aan de diversiteit van het HPV genoom.

In **hoofdstuk 7** hebben we HPV16 varianten bekeken van vrouwen die behandeld zijn voor CIN2/3 laesies en die op controlemomenten wederom een HPV16 positieve recidive CIN (rCIN) bleken te hebben. Gewoonlijk klaart de meerderheid (90–95%) van de vrouwen die behandeling (LLETZ of LEEP) ondergaan na CIN2/3 detectie het eventuele restant van de infectie op eigen kracht. In een minderheid van de gevallen (5–10%) is de behandelde vrouw niet in staat de infectie te klaren en ontwikkelt zij rCIN. Sequentiebepaling van zowel de HPV16 variant in de CIN2/3 als de rCIN één tot twee jaar later leverde in beide gevallen (nagenoeg) dezelfde virusvariant op. De aanwezigheid van dezelfde HPV16 virusvariant lijkt erop te wijzen dat voor deze minderheidsgroep, de behandeling niet effectief is geweest, of dat deze vrouwen niet in staat zijn het postoperatieve restant van de infectie op eigen kracht te klaren.

Kruisbescherming van HPV vaccinatie tegen niet-vaccin types wordt momenteel breed onderzocht en hoewel er consensus bestaat dat kruisbescherming bestaat, zijn er wel verschillen tussen studies die dit effect bestuderen. In **hoofdstuk 8** van dit proefschrift correleren we HPV referentiesequenties aan HPV vaccin effectiviteitsdata. We tonen aan dat de fylogenetische afstand van een HPV type ten opzichte van de HPV16 of HPV18 eiwitsequentie sterk gecorreleerd is met de te verwachten bescherming tegen dat HPV type. Kruisbescherming bleek sterker tegen HPV types die nauw verwant zijn aan de HPV types waartegen het bivalente vaccin beschermt. Daarentegen bleek kruisbescherming tegen HPV types die niet verwant zijn aan vaccintypes onwaarschijnlijk. Deze bevinding kan mogelijk helpen de verschillen tussen eerdere studies te verklaren.

De studies die in dit proefschrift beschreven zijn hebben geresulteerd in nieuwe technieken om de aanwezigheid van persisterende HPV infecties te monitoren in gevaccineerde vrouwen. Aan de hand van viral load bepalingen op populatieniveau konden we op een andere manier bevestigen dat het bivalente HPV vaccin in staat is persistente HPV16 en HPV18 infecties te voorkomen en dus waarschijnlijk bijzonder goed beschermt tegen baarmoederhalskanker.

Door middel van meerdere methoden van sequentiebepaling hebben we aan kunnen tonen dat het HPV genoom sterk behouden blijft over tijd, terwijl het toch ook zeer divers is tussen mensen en zelfs binnen mensen. Verder hebben we laten zien dat variantbepaling van HPV16 en HPV18 nuttig kan blijken om herinfecties met verschillende varianten van hetzelfde HPV type uit te sluiten in doorbraakinfecties onder gevaccineerde vrouwen.

About the author

Pascal van der Weele was born on November 11, 1986 in Breda, the Netherlands. After finishing high school, he studied Biology at Utrecht University, where he took courses on a broad range of biological subjects. Here, he developed a strong interest towards molecular biology. After obtaining his Bachelor's degree, he obtained a Master's degree in Molecular and Cellular Life Sciences at Utrecht University as well, with a focus on molecular genetics. He did a one-year internship in plant molecular genetics, under the supervision of dr. René Benjamins. After this internship, he chose to continue doing research with a more direct relation to humans and society. His second internship was performed at the National Institute for Public Health and the Environment (RIVM) on the molecular biology of human papillomavirus (HPV) infection under the supervision of dr. Audrey King. Here he discovered his interest in virology and public health, combined with molecular biological research. After this internship, he proceeded to do his doctoral research on HPV at the RIVM as well, in collaboration with the Free University Medical Center (VUmc), where he worked under the joint supervision of prof. dr. Chris Meijer and dr. Audrey King. During his doctoral research, he was able to combine molecular virological research with public health and clinical research topics.

List of publications

van der Weele P., van Logchem E., Wolffs P., van den Broek I., Feltkamp M., de Melker H., *et al.* Correlation between viral load, multiplicity of infection, and persistence of HPV16 and HPV18 infection in a Dutch cohort of young women. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology. 2016;83:6-11.

van der Weele P., Meijer C.J.L.M., King A.J. Whole-Genome Sequencing and Variant Analysis of Human Papillomavirus 16 Infections. Journal of virology. 2017;91(19).

van der Weele P., Meijer C.J.L.M., King A.J. High Whole-Genome Sequence Diversity of Human Papillomavirus Type 18 Isolates. Viruses. 2018;10(2):68.

Marra E., King A.J., van Logchem E., **van der Weele P.**, Mooij S.H., Heijman T., *et al.* Anal HPV 16 and 18 viral load: A comparison between HIV-negative and -positive MSM and association with persistence. Journal of medical virology. 2018;90(1):76-83.

van der Weele P., Meijer C.J.L.M., King A.J. High Whole-Genome Sequence Diversity of Human Papillomavirus Type 18 Isolates. Viruses. 2018;10(2).

Bogaards J.A., **van der Weele P.**, Woestenberg P.J., van Benthem B.H.B., King A.J. Bivalent HPV Vaccine Effectiveness Correlates with Phylogenetic Distance from Hpv Vaccine Types 16 And 18. The Journal of infectious diseases. 2019.

van der Weele P., King A.J., Meijer C.J.L.M., Steenbergen R.D.M. HPV16 variant analysis in primary and recurrent CIN2/3 lesions demonstrates presence of the same consensus variant. Papillomavirus Res. 2019;7:168-72.

van der Weele P., Breeuwsma M., Donken R., van Logchem E., van Marm-Wattimena N., de Melker H., *et al.* Effect of the bivalent HPV vaccine on viral load of vaccine and non-vaccine HPV types in incident clearing and persistent infections in young Dutch females. PloS one. 2019;14(3).

Lagström S., **van der Weele P.**, Rounge T.B., Christiansen I.K., King A.J., Ambur O.H. HPV16 whole genome minority variants in persistent infections from young Dutch women. J Clin Virol. 2019 Aug 12;119:24-30.

Dankwoord

Dr. Audrey King, Audrey; bedankt voor het vertrouwen en de prettige samenwerking de afgelopen jaren. Via Frank ben ik bij jou terecht gekomen voor mijn masterstage en hier heb ik geen dag spijt van gehad! Ik heb het erg naar mijn zin gehad op het RIVM en dankzij jou ben ik doelgericht en pragmatisch na gaan denken over onderzoeksvragen en het onderzoek zelf. Je hebt me altijd gestimuleerd het beste uit mezelf en het onderzoek te halen.

Prof. dr. Chris Meijer, Chris; jou wil ik ook graag hartelijk danken voor je begeleiding gedurende mijn promotie. Volgens mij heb je in het begin aardig wat moeten schaven aan mijn klinische kennis en er moet vast nog een hoop aan gebeuren, maar wat heb ik veel van jou geleerd zeg. Als ik weer een of ander idee had, was het altijd binnen een kwartier: "Ja leuk, maar wat is nou de klinische relevantie?" en dan kon ik me weer achter mijn oren gaan krabben. Dankzij jou ben ik kritischer gaan denken en vooral breder dan alleen maar moleculair. Ik vind het erg prettig dat ik je altijd kon bellen, of het nou over het werk ging of niet en dat je altijd achter me hebt gestaan.

Mijn directe HPV collega's met wie ik door de jaren heen heb samengewerkt: Suzan, Elske, Naomi, Jan, Jason, Kahren en Tim. Ik vond het altijd erg gezellig om met jullie samen te werken. Zowel op persoonlijk als op professioneel vlak kon ik het goed met jullie vinden. Vooral de vele bakexperimenten van Suzan en de slechte Whatsapp grappen van Tim zullen me lang bij blijven. Kahren, jou wil ik graag heel veel succes wensen met jouw promotietraject. Ik hoop dat je al lang niet meer gezien wordt als 'Pascal 2.0' en dat je een mooie eigen draai kunt geven aan het onderzoek. Ik hoor graag hoe jouw boekje vordert.

Martijn, wat heb jij een berg werk verzet zeg! Zonder jou zat ik waarschijnlijk nu nog op het lab qPCR's in te zetten. Ik denk dat ik het erg heb getroffen met jou als stagiair. Je pikte alles ontzettend snel op en kwam met leuke ideeën voor nieuw onderzoek of alternatieve methodes. Ik kon er altijd erg om lachen als we het ergens niet over eens waren, we waren dan allebei te koppig om water bij de wijn te doen. Als het om de methode of interpretatie van de resultaten ging, moesten er altijd heel wat feiten bijgesleurd worden voordat we de discussie los konden laten. Uiteindelijk heeft je stage geresulteerd in een belachelijk hoog punt, maar je hebt het zeker verdiend. Je ambitie was duidelijk zichtbaar en ik ben benieuwd naar wat je nog allemaal gaat bereiken.

De 'cool people' Sigrid, Seta, Natalie, Brenda, Reina en Janko natuurlijk. Dank voor de leuke gesprekken, etentjes en discussies op kantoor en daarbuiten!

Dennis en Sam, jullie wil ik graag bedanken voor al het geduld dat jullie konden opbrengen bij het beantwoorden van mijn (vele) noob bioinformatica vragen. Ik heb er ontzettend veel van geleerd, vooral dat spaties in python niet hetzelfde zijn als tabs!

Met veel plezier heb ik tijdens mijn tweede masterstage en tijdens mijn promotietraject deel uitgemaakt van IDS-VIR. Ik wil graag alle collega's bedanken voor de afgelopen jaren. Rob, inmiddels zit je bij de immunologie, maar ik heb bijzonder veel aan je input gehad. Het kostte soms wat moeite om te destilleren wat ik concreet kon doen om je een antwoord te geven op je vraag, maar je vragen waren altijd discipline overstijgend en kwamen voort uit oprechte interesse.

Mijn mede HPV onderzoekers van de epidemiologie; Hester, Robine, Hans, Petra, Birgit, Irene en Joske. Ontzettend bedankt voor alle input en gezellige etentjes en congressen. Het was inhoudelijk heel stimulerend om regelmatig samen te zitten over de verschillende onderzoeken. Robine, ik hoop dat je het naar je zin hebt in Canada. Ik vond het altijd fijn om met je te sparren over het onderzoek. Ook was het erg prettig dat je altijd de tijd nam om de soms complexe, maar ook de minder complexe epidemiologische analyses aan me uit te leggen.

De collega's van de immunologie; Fiona en Hella. Net als met de collega's van EPI was het erg fijn om met jullie te sparren over de onderwerpen van HPV onderzoek waar ik minder van wist.

Herman, Trine, Irene and Sonja, thank you very much for welcoming me in Norway and for the great collaboration opportunity.

De collega's van het VUmc en het CCA wil ik graag bedanken voor de kritische discussies en gezellige congressen. Peter, helaas ben je ons veel te vroeg ontvallen. Ik kon altijd erg om je lachen tijdens de meetings. Jij, Daniëlle en Renske waren altijd bereid mijn projecten te bespreken en input te geven vanuit de klinisch moleculaire hoek.

Mijn paranimfen, Michelle en Rens.

Michelle, Miepje, ik ben er trots op hoe je jouw leven met Tim in Australië invulling geeft. Het zit je niet altijd mee, maar je geeft nooit op en kachelt stug verder. Ik hoop heel erg dat het je de komende tijd lukt om een baan te vinden waarin je echt tot je recht kan komen. That goes for you too Tim. Just ask Miep to translate the previous bit, unless you have finally decided to push through with learning Dutch! Thanks for letting me borrow Michelle for this and I'll be sure to send a nice bottle of gin and some 'boring' cake back downunder.

Rens, ik vind het erg fijn dat je mijn paranimf wilt zijn. Jaren geleden hadden we al grappend afgesproken dat als jij ooit zou willen promoveren, dat ik jouw paranimf zou zijn en vice versa. Jij bent dan misschien niet gaan promoveren, maar ik vind het leuk dat ik toch de daad bij het woord kan voegen. Ik weet zeker dat de rest van de 'iets generieks' blij is dat deze promotie nu eindelijk achter de rug is. Al betekent dat natuurlijk niet dat we het niet meer over allerlei farmaceutische middeltjes en moleculaire technieken kunnen hebben.

Mijn favoriete zooitje ongeregeld uit Etten-Leur, de groep die tegenwoordig door het leven gaat als 'iets generieks'. Sorry voor de vele saaie discussies over farmacie en biologie met Rens. Nu is het klaar hoor, wat betreft dit onderzoek dan…

De vriendjes en vriendinnetjes van 'Warande 116, de' en natuurlijk Merel en Judith die ik tijdens mijn studie heb leren kennen. Bedankt voor de vele etentjes, gezellige avonden en foute feestjes.

En natuurlijk de NVH. Jullie kreeg ik min of meer cadeau bij Carmen. Voor een stel Rotterdammers zijn jullie best oké! Ik kom nu zelfs met plezier jullie kant op.

Mijn familie, waarvan iedereen altijd geïnteresseerd was in het reilen en zeilen van mijn onderzoek. Ik vind het erg fijn dat jullie er vandaag bij willen zijn. Opa, helaas kan jij er niet meer bij zijn, maar ik weet zeker dat je dit fantastisch zou vinden.

Mam, jou wil ik graag bedanken voor het feit dat je altijd in me bent blijven geloven, zelfs wanneer het van mij allemaal niet meer hoefde. Als ik weer eens te lui of ongemotiveerd was om te leren en studeren, sleepte jij me er altijd doorheen. Het heeft op zich laten wachten, maar uiteindelijk heb toch datgene gevonden waar ik wel enthousiast voor kon worden, de moleculaire biologie. Ik heb je vaak over mijn onderzoek verteld, en over de grote lijnen konden we goed met elkaar praten. Wanneer ik voor de zoveelste keer in te veel detail trad liet je mij altijd doorbrabbelen. Je werd denk ik vooral enthousiast van mijn overtuiging! Geen flauw idee waar ik was geweest zonder jouw (soms dwingende) peptalks!

Lieve Carmen, lieve schat, jij hebt altijd met liefde en plezier mijn frustraties en successen aangehoord gedurende mijn promotie. Je hebt je verdiept in mijn onderzoek en inmiddels kan ik met gemak met je praten over allerlei moleculaire onderwerpen en allerhande andere dingen die komen kijken bij een promotietraject. Samen met Radja was jij de afgelopen jaren mijn steun en toeverlaat.